Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2478957 C 2013/07/02

(11)(21) 2 478 957

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2003/03/18
(87) Date publication PCT/PCT Publication Date: 2003/10/02
(45) Date de délivrance/Issue Date: 2013/07/02
(85) Entrée phase nationale/National Entry: 2004/09/09
(86) N° demande PCT/PCT Application No.: US 2003/008468
(87) N° publication PCT/PCT Publication No.: 2003/080647
(30) Priorités/Priorities: 2002/03/19 (US60/365,202);
2003/03/18 (US10/391,363)

(51) Cl.Int./Int.Cl. *C12N 15/54* (2006.01),
A01H 5/00 (2006.01), *A01H 5/10* (2006.01),
C07K 14/415 (2006.01), *C07K 16/16* (2006.01),
C07K 16/40 (2006.01), *C12N 15/29* (2006.01),
C12N 15/82 (2006.01), *C12N 9/10* (2006.01),
C12P 17/06 (2006.01)

(72) Inventeurs/Inventors:
VALENTIN, HENRY E., US;
VENKATESH, TYAMAGONDLU V., US;
KARUNANANDAA, BALASULOJINI, US

(73) Propriétaire/Owner:
MONSANTO TECHNOLOGY, LLC, US

(74) Agent: MCFADDEN, FINCHAM INC.

(54) Titre : ACIDES NUCLEIQUES ET POLYPEPTIDES SE RAPPORTANT A L'ENZYME HOMOGENTISATE PRENYL TRANSFERASE ("HPT"), ET UTILISATIONS CORRESPONDANTES

(54) Title: HOMOGENTISATE PRENYL TRANSFERASE ("HPT") NUCLEIC ACIDS AND POLYPEPTIDES, AND USES THEREOF

(57) Abrégé/Abstract:
The present invention is in the field of plant genetics and biochemistry. More specifically, the present invention relates to genes and polypeptides associated with the tocopherol biosynthesis pathway, namely those encoding homogentisate prenyl transferase activity, and uses thereof.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 October 2003 (02.10.2003)

PCT

(10) International Publication Number
WO 03/080647 A2

(51) International Patent Classification[7]: C07K

(21) International Application Number: PCT/US03/08468

(22) International Filing Date: 18 March 2003 (18.03.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/365,202	19 March 2002 (19.03.2002)	US
10/391,363	18 March 2003 (18.03.2003)	US

(71) Applicant *(for all designated States except US)*: **MONSANTO TECHNOLOGY, LLC** [US/US]; 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **VALENTIN, Henry, E.** [DE/US]; 873 M Foxsprings Drive, Chesterfield, MO 63017 (US). **VENKATESH, Tyamagondlu, V.** [IN/US]; 11747 Suummerhaven Drive, St. Louis, MO 63146 (US). **KARUNANANDAA, Balasulojini** [LK/US]; 724 Bellerive Manor, Creve Coeur, MO 63141 (US).

(74) Agents: **SILVERT, Donald, J.** et al.; Renessen LLC, 3000 Lakeside Drive, Suite 300 South, Bannockburn, IL 60015 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for all designations*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*
— *of inventorship (Rule 4.17(iv)) for US only*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/080647 A2

(54) Title: HOMOGENTISATE PRENYL TRANSFERASE ("HPT") NUCLEIC ACIDS AND POLYPEPTIDES, AND USES THEREOF

(57) Abstract: The present invention is in the field of plant genetics and biochemistry. More specifically, the present invention relates to genes and polypeptides associated with the tocopherol biosynthesis pathway, namely those encoding homogentisate prenyl transferase activity, and uses thereof.

HOMOGENTISATE PRENYL TRANSFERASE ("HPT") NUCLEIC ACIDS AND POLYPEPTIDES, AND USES THEREOF

The present invention is in the field of plant genetics and biochemistry. More specifically, the present invention relates to genes and polypeptides associated with the tocopherol biosynthesis pathway, namely those encoding homogentisate prenyl transferase activity, and uses thereof.

Isoprenoids are ubiquitous compounds found in all living organisms. Plants synthesize a diverse array of greater than 22,000 isoprenoids (Connolly and Hill, *Dictionary of Terpenoids,* Chapman and Hall, New York, NY (1992)). In plants, isoprenoids play essential roles in particular cell functions such as production of sterols, contributing to eukaryotic membrane architecture, acyclic polyprenoids found in the side chain of ubiquinone and plastoquinone, growth regulators like abscisic acid, gibberellins, brassinosteroids or the photosynthetic pigments chlorophylls and carotenoids. Although the physiological role of other plant isoprenoids is less evident, like that of the vast array of secondary metabolites, some are known to play key roles mediating the adaptative responses to different environmental challenges. In spite of the remarkable diversity of structure and function, all isoprenoids originate from a single metabolic precursor, isopentenyl diphosphate (IPP) (Wright, (1961) *Annu. Rev. Biochem.*, 20:525-548; and Spurgeon and Porter, In: *Biosynthesis of Isoprenoid Compounds*, Porter and Spurgeon (eds.) John Wiley, NY, Vol. 1, pp. 1-46 (1981)).

A number of unique and interconnected biochemical pathways derived from the isoprenoid pathway leading to secondary metabolites, including tocopherols, exist in chloroplasts of higher plants. Tocopherols not only perform vital functions in plants, but are also important from mammalian nutritional perspectives. In plastids, tocopherols account for up to 40% of the total quinone pool. Tocopherols are an important component of mammalian diets. Epidemiological evidence indicates that tocopherol supplementation can result in decreased risk for cardiovascular disease and cancer, can aid in immune function, and is associated with prevention or retardation of a number of degenerative disease processes in humans (Traber and Sies, *Annu. Rev. Nutr.*, 16:321-347 (1996)). Tocopherol functions, in part, by stabilizing the lipid bilayer of biological membranes (Skrypin and Kagan, *Biochim. Biophys. Acta*, 815:209 (1995); Kagan, *N.Y. Acad. Sci.*, p. 121 (1989); Gomez-Fernandez *et al.*, *Ann. N.Y. Acad. Sci.*, p. 109 (1989)), reducing polyunsaturated fatty acid (PUFA) free

radicals generated by lipid oxidation (Fukuzawa *et al.*, *Lipids*, 17:511-513 (1982)), and scavenging oxygen free radicals, lipid peroxy radicals and singlet oxygen species (Diplock *et al.*, *Ann. N Y Acad. Sci.*, 570:72 (1989); Fryer, *Plant Cell Environ.*, 15(4):381-392 (1992)).

The compound α-tocopherol, which is often referred to as vitamin E, belongs to a class of lipid-soluble antioxidants that includes α, β, γ, and δ-tocopherols and α, β, γ, and δ-tocotrienols. Although α, β, γ, and δ-tocopherols and α, β, γ, and δ-tocotrienols are sometimes referred to collectively as "vitamin E", vitamin E is more appropriately defined chemically as α-tocopherol. Vitamin E, or α-tocopherol, is significant for human health, in part because it is readily absorbed and retained by the body, and therefore has a higher degree of bioactivity than other tocopherol species (Traber and Sies, *Annu. Rev. Nutr.*, 16:321-347 (1996)). Other tocopherols, however, such as β, γ, and δ-tocopherols also have significant health and nutritional benefits.

Tocopherols are primarily synthesized only by plants and certain other photosynthetic organisms, including cyanobacteria. As a result, mammalian dietary tocopherols are obtained almost exclusively from these sources. Plant tissues vary considerably in total tocopherol content and tocopherol composition, with α-tocopherol the predominant tocopherol species found in green, photosynthetic plant tissues. Leaf tissue can contain from 10-50 μg of total tocopherols per gram fresh weight, but most of the world's major staple crops (*e.g.*, rice, corn, wheat, potato) produce low to extremely low levels of total tocopherols, of which only a small percentage is α-tocopherol (Hess, Vitamin E, α-tocopherol, In: *Antioxidants in Higher Plants*, R. Alscher and J. Hess, (eds.), CRC Press, Boca Raton., pp. 111-134 (1993)). Oil seed crops generally contain much higher levels of total tocopherols, but α-tocopherol is present only as a minor component in most oilseeds (Taylor and Barnes, *Chemy Ind.*, Oct:722-726 (1981)).

The recommended daily dietary intake of 15-30 mg of vitamin E is quite difficult to achieve from the average American diet. For example, it would take over 750 grams of spinach leaves, in which α-tocopherol comprises 60% of total tocopherols, or 200-400 grams of soybean oil to satisfy this recommended daily vitamin E intake. While it is possible to augment the diet with supplements, most of these supplements contain primarily synthetic vitamin E, having eight stereoisomers, whereas natural vitamin E is predominantly composed of only a single isomer. Furthermore, supplements tend to be relatively expensive, and the general population is disinclined to take vitamin supplements on a regular basis. Therefore,

there is a need in the art for compositions and methods that either increase the total tocopherol production or increase the relative percentage of α-tocopherol produced by plants.

In addition to the health benefits of tocopherols, increased α-tocopherol levels in crops have been associated with enhanced stability and extended shelf life of plant products (Peterson, *Cereal-Chem.*, 72(1):21-24 (1995); Ball, *Fat-soluble vitamin assays in food analysis. A comprehensive review,* London, Elsevier Science Publishers Ltd. (1988)). Further, tocopherol supplementation of swine, beef, and poultry feeds has been shown to significantly increase meat quality and extend the shelf life of post-processed meat products by retarding post-processing lipid oxidation, which contributes to the undesirable flavor components (Sante and Lacourt, *J. Sci. Food Agric.*, 65(4):503-507 (1994); Buckley *et al.*, *J. of Animal Science*, 73:3122-3130 (1995)).

TOCOPHEROL BIOSYNTHESIS

The plastids of higher plants exhibit interconnected biochemical pathways leading to secondary metabolites including tocopherols. The tocopherol biosynthetic pathway in higher plants involves condensation of homogentisic acid and phytylpyrophosphate to form 2-methylphytylplastoquinol (Fiedler *et al.*, *Planta*, 155:511-515 (1982); Soll *et al.*, *Arch. Biochem. Biophys.*, 204:544-550 (1980); Marshall *et al.*, *Phytochem.*, 24:1705-1711 (1985)). This plant tocopherol pathway can be divided into four parts: 1) synthesis of homogentisic acid (HGA), which contributes to the aromatic ring of tocopherol; 2) synthesis of phytylpyrophosphate, which contributes to the side chain of tocopherol; 3) joining of HGA and phytylpyrophosphate via a homogentisate prenyl transferase followed by a subsequent cyclization; and 4) S-adenosyl methionine dependent methylation of an aromatic ring, which affects the relative abundance of each of the tocopherol species. *See* Figure 1.

Various genes and their encoded proteins that are involved in tocopherol biosynthesis include those listed in the table below:

Gene ID	Enzyme name
tyrA	Bifunctional prephenate dehydrogenase
slr1736	Homogentisate prenyl transferase from Synechocystis
ATPT2	Homogentisate prenyl transferase from Arabidopsis thaliana

Gene ID	Enzyme name
DXS	1-Deoxyxylulose-5- phosphate synthase
DXR	1-Deoxyxylulose-5- phosphate reductoisomerase
GGPPS	Geranylgeranyl pyrophosphate synthase
HPPD	p-Hydroxyphenylpyruvate dioxygenase
AANT1	Adenylate transporter
slr1737	Tocopherol cyclase
IDI	Isopentenyl diphosphate isomerase
GGH	Geranylgeranyl diphosphate reductase
GMT	Gamma Methyl Transferase
tMT2	Tocopherol methyl transferase 2
MT1	Methyl transferase 1
gcpE	(E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase

The "Gene IDs" given in the table above identify the gene associated with the listed enzyme. Any of the Gene IDs listed in the table appearing herein in the present disclosure refer to the gene encoding the enzyme with which the Gene ID is associated in the table.

As used herein, *HPT*, *HPT2*, *PPT*, *slr1736*, and *ATPT2* each refer to proteins or genes encoding proteins that have the same activity.

SYNTHESIS OF HOMOGENTISIC ACID

Homogentisic acid is the common precursor to both tocopherols and plastoquinones. In at least some bacteria the synthesis of homogentisic acid is reported to occur via the conversion of chorismate to prephenate and then to p-hydroxyphenylpyruvate via a bifunctional prephenate dehydrogenase. Examples of bifunctional bacterial prephenate dehydrogenase enzymes include the proteins encoded by the *tyrA* genes of *Erwinia herbicola* and *Escherichia coli*. The *tyrA* gene product catalyzes the production of prephenate from chorismate, as well as the subsequent dehydrogenation of prephenate to form p-hydroxyphenylpyruvate (p-HPP), the immediate precursor to homogentisic acid. p-HPP is then converted to homogentisic acid by hydroxyphenylpyruvate dioxygenase (HPPD). In contrast, plants are believed to lack prephenate dehydrogenase activity, and it is generally believed that the synthesis of homogentisic acid from chorismate occurs via the synthesis and conversion of the intermediate arogenate. Since pathways involved in homogentisic acid

synthesis are also responsible for tyrosine formation, any alterations in these pathways can also result in the alteration in tyrosine synthesis and the synthesis of other aromatic amino acids.

SYNTHESIS OF PHYTYLPYROPHOSPHATE

Tocopherols are a member of the class of compounds referred to as the isoprenoids. Other isoprenoids include carotenoids, gibberellins, terpenes, chlorophyll and abscisic acid. A central intermediate in the production of isoprenoids is isopentenyl diphosphate (IPP). Cytoplasmic and plastid-based pathways to generate IPP have been reported. The cytoplasmic based pathway involves the enzymes acetoacetyl CoA thiolase, HMGCoA synthase, HMGCoA reductase, mevalonate kinase, phosphomevalonate kinase, and mevalonate pyrophosphate decarboxylase.

Recently, evidence for the existence of an alternative, plastid based, isoprenoid biosynthetic pathway emerged from studies in the research groups of Rohmer and Arigoni (Eisenreich *et al.*, *Chem. Bio.*, 5:R221-R233 (1998); Rohmer, *Prog. Drug. Res.*, 50:135-154 (1998); Rohmer, *Comprehensive Natural Products Chemistry*, Vol. 2, pp. 45-68, Barton and Nakanishi (eds.), Pergamon Press, Oxford, England (1999)), who found that the isotope labeling patterns observed in studies on certain eubacterial and plant terpenoids could not be explained in terms of the mevalonate pathway. Arigoni and coworkers subsequently showed that 1-deoxyxylulose, or a derivative thereof, serves as an intermediate of the novel pathway, now referred to as the MEP pathway (Rohmer *et al.*, *Biochem. J.*, 295:517-524 (1993); Schwarz, Ph.D. thesis, Eidgenössiche Technische Hochschule, Zurich, Switzerland (1994)). Recent studies showed the formation of 1-deoxyxylulose 5-phosphate (Broers, Ph.D. thesis, Eidgenössiche Technische Hochschule, Zurich, Switzerland (1994)) from one molecule each of glyceraldehyde 3-phosphate (Rohmer, *Comprehensive Natural Products Chemistry*, Vol. 2, pp. 45-68, Barton and Nakanishi (eds.), Pergamon Press, Oxford, England (1999)) and pyruvate (Eisenreich *et al.*, *Chem. Biol.*, 5:R223-R233 (1998); Schwarz *supra*; Rohmer *et al.*, *J. Am. Chem. Soc.*, 118:2564-2566 (1996); and Sprenger *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 94:12857-12862 (1997)) by an enzyme encoded by the *dxs* gene (Lois *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 95:2105-2110 (1997); and Lange *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 95:2100-2104 (1998)). 1-Deoxyxylulose 5-phosphate can be further converted into 2-*C*-methylerythritol 4-phosphate (Arigoni *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 94:10600-10605 (1997)) by a reductoisomerase encoded by the *dxr* gene (Bouvier *et al.*, *Plant*

Physiol, 117:1421-1431 (1998); and Rohdich *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 96:11758-11763 (1999)).

Reported genes in the MEP pathway also include *ygbP,* which catalyzes the conversion of 2-*C*-methylerythritol 4-phosphate into its respective cytidyl pyrophosphate derivative and *ygbB,* which catalyzes the conversion of 4-phosphocytidyl-2-*C*-methyl-D-erythritol into 2-*C*-methyl-D-erythritol, 3, 4-cyclophosphate. These genes are tightly linked on the *E. coli* genome (Herz *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 97(6):2485-2490 (2000)).

Once IPP is formed by the MEP pathway, it is converted to GGDP by GGPDP synthase, and then to phytylpyrophosphate, which is the central constituent of the tocopherol side chain.

COMBINATION AND CYCLIZATION

Homogentisic acid is combined with either phytyl-pyrophosphate or solanyl-pyrophosphate by homogentisate prenyl transferase (*HPT*) forming 2-methylphytyl plastoquinol or 2-methylsolanyl plastoquinol, respectively. 2-methylsolanyl plastoquinol is a precursor to the biosynthesis of plastoquinones, while 2-methylphytyl plastoquinol is ultimately converted to tocopherol.

METHYLATION OF THE AROMATIC RING

The major structural difference between each of the tocopherol subtypes is the position of the methyl groups around the phenyl ring. Both 2-methylphytyl plastoquinol and 2-methylsolanyl plastoquinol serve as substrates for the plant enzyme 2-methylphytylplastoquinol/2-methylsolanylplastoquinol methyltransferase (Tocopherol Methyl Transferase 2; Methyl Transferase 2; MT2; tMT2), which is capable of methylating a tocopherol precursor. Subsequent methylation at the 5 position of γ-tocopherol by γ-tocopherol methyl-transferase (GMT) generates the biologically active α-tocopherol.

Some plants e.g. soy produce substantial amounts of delta and subsequently beta-tocopherol in their seed. The formation of δ-tocopherol or β-tocopherol can be prevented by the overexpression of tMT2, resulting in the methylation of the δ-tocopherol precursor, 2-methyl phytyl plastoquinone to form 2, 3-dimethyl-5-phytyl plastoquinone followed by cyclization with tocopherol cyclase to form γ-tocopherol and a subsequent methylation by GMT to form α-tocopherol. In a possible alternative pathway, β-tocopherol is directly converted to α-tocopherol by tMT2 via the methylation of the 3 position (*see*, for example, *Biochemical Society Transactions,* 11:504-510 (1983); *Introduction to Plant Biochemistry*, 2nd

edition, Chapter 11 (1983); *Vitamin Hormone,* 29:153-200 (1971); *Biochemical Journal,* 109:577 (1968); and, *Biochemical and Biophysical Research Communication,* 28(3):295 (1967)). Since all potential mechanisms for the generation of α-tocopherol involve catalysis by tMT2, plants that are deficient in this activity accumulate δ-tocopherol and β-tocopherol. Plants which have increased tMT2 activity tend to accumulate γ-tocopherol and α-tocopherol. Since there is limited GMT activity in the seeds of many plants, these plants tend to accumulate γ-tocopherol.

There is a need in the art for nucleic acid molecules encoding enzymes involved in tocopherol biosysnthesis, as well as related enzymes and antibodies for the enhancement or alteration of tocopherol production in plants. There is a further need for transgenic organisms expressing those nucleic acid molecules involved in tocopherol biosynthesis, which are capable of nutritionally enhancing food and feed sources.

SUMMARY OF THE INVENTION

The present invention includes and provides a substantially purified nucleic acid molecule encoding an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 57-58, and 90.

The present invention includes and provides a substantially purified polypeptide molecule comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 57-58, and 90.

The present invention includes and provides an antibody capable of specifically binding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 57-58, and 90.

The present invention includes and provides a substantially purified nucleic acid molecule encoding a polypeptide having homogentisate prenyl transferase activity comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 43 and 44.

The present invention includes and provides a substantially purified polypeptide having homogentisate prenyl transferase activity comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 43 and 44.

The present invention includes and provides a transformed plant comprising an introduced nucleic acid molecule encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof.

The present invention includes and provides a transformed plant comprising an introduced first nucleic acid molecule encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof, and an introduced second nucleic acid molecule encoding an enzyme selected from the group consisting of *tyrA, prephenate dehyrogenase, tocopherol cyclase, dxs, dxr, GMT, MT1, tMT2, GCPE, GGPPS, HPPD, AANT1, IDI, GGH*, and complements thereof.

The present invention includes and provides a transformed plant comprising a nucleic acid molecule comprising an introduced promoter region which functions in plant cells to cause the production of an mRNA molecule, wherein said introduced promoter region is linked to a transcribed nucleic acid molecule having a transcribed strand and a non-transcribed strand, wherein said transcribed strand is complementary to a nucleic acid molecule encoding a polypeptide selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and wherein said transcribed nucleic acid molecule is linked to a 3' non-translated sequence that functions in the plant cells to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA sequence.

The present invention includes and provides a method of producing a plant having a seed with an increased total tocopherol level comprising: (A) transforming said plant with an introduced nucleic acid molecule encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90; and (B) growing said transformed plant.

The present invention includes and provides a method of producing a plant having a seed with an increased total tocopherol level comprising: (A) transforming said plant with an introduced first nucleic acid molecule, wherein said first nucleic acid molecule encodes a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and an introduced second nucleic acid molecule encoding an enzyme selected from the group consisting of *tyrA, prephenate dehydrogenase, tocopherol cyclase, dxs, dxr, GMT, MT1, tMT2, GGPPS, GCPE, HPPD, AANT1, IDI, GGH*, and complements thereof; and (B) growing said transformed plant.

The present invention includes and provides a seed derived from a transformed plant comprising an introduced nucleic acid molecule encoding a polypeptide comprising an amino

acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90.

The present invention includes and provides a seed derived from a transformed plant comprising an introduced first nucleic acid molecule encoding an introduced polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43, 44, 57, 58, and 90 and an introduced second nucleic acid encoding an enzyme selected from the group consisting of *tyrA, prephenate dehydrogenase, tocopherol cyclase, dxs, dxr, GMT, MT1, GCPE, tMT2, GGPPS, HPPD, AANT1, IDI, GGH*, and complements thereof.

The present invention includes and provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95 wherein said amino acid sequence is not derived from a nucleic acid molecule that is derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa*, Trichodesmium erythraeum, Chloroflexus aurantiacus, wheat, leek, canola, cotton, or tomato. The present invention includes and provides said substantially purified polypeptide wherein more than one more amino acid sequence is selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

The present invention includes and provides a substantially purified nucleic acid molecule encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95 wherein said nucleic acid molecule is not derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato. The present invention includes and provides said nucleic acid molecule wherein the polypeptide further comprises more than one amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

The present invention includes and provides a substantially purified nucleic acid molecule encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95 wherein said nucleic acid molecule is not derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Sulfolobus, Aeropyum, Trichodesmium erythraeum, Chloroflexus aurantiacus,* sorghum, wheat, tomato, or leek. The present invention includes and provides said nucleic acid molecule wherein the polypeptide further comprises more than

one amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49 and 92-95.

The present invention includes and provides a plant transformed with a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95 wherein said nucleic acid molecule is not derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Sulfolobus, Aeropyum, Trichodesmium erythraeum, Chloroflexus aurantiacus,* sorghum, wheat, tomato, or leek. The present invention includes and provides said nucleic acid molecule wherein the polypeptide further comprises more than one amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

The present invention includes and provides a substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95 wherein said polypeptide does not comprise any of the amino acid sequences set forth in sequence listings in WO 00/68393; WO 00/63391; or WO 02/33060; and does not comprise SEQ ID NOs: 1-11, 43-45, 57-58, 61-62, or 90 from the present application.

The present invention includes and provides a substantially purified polypeptide comprising more than onean amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

The present invention includes and provides a substantially purified nucleic acid molecule encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95 wherein said nucleic acid molecule does not comprise any of the nucleic acid sequences set forth in sequence listings in WO 00/68393; WO 00/63391; WO 01/62781; or WO 02/33060; and does not comprise SEQ ID NOs: 27-36, 59-60, 88-89, and 91 from the present application, or the gene with Genebank Accession Nos. AI 897027 or AW 563431 The present invention includes and provides said nucleic acid molecule wherein the polypeptide further comprises more than one amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

The present invention includes and provides a plant transformed with a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95 wherein said nucleic acid molecule does not comprise any of the nucleic acid sequences set forth in sequence listings in WO 00/68393; WO 00/63391; WO 01/62781; or WO 02/33060; and does not comprise SEQ ID NOs: 27-36; 59-60, 88-89, and 91 from the present application, or the gene with Genebank Accession Nos. AI 897027 or AW 563431. The present invention includes and provides said nucleic acid molecule wherein the polypeptide further comprises more than one amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

The present invention includes and provides a substantially purified nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 31, 34-36, 59-60, and 91.

The present invention includes and provides for homogentisate prenyl transferases discovered using one or more of the alignments of Figures 2a-2c, 3a-3c, 24a-24b, 25a-25b, 33a-33c, 34a-34b, 35a-35b and 36.

DESCRIPTION OF THE NUCLEIC AND AMINO ACID SEQUENCES

SEQ ID NO: 1 sets forth a *Nostoc punctiforme* homogentisate prenyl transferase polypeptide.

SEQ ID NO: 2 sets forth an *Anabaena* homogentisate prenyl transferase polypeptide.

SEQ ID NO: 3 sets forth a *Synechocystis* homogentisate prenyl transferase polypeptide.

SEQ ID NO: 4 sets forth a *Zea mays* homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NO: 5 sets forth a *Glycine max* homogentisate prenyl transferase polypeptide (HPT1-2).

SEQ ID NO: 6 sets forth a *Glycine max* homogentisate prenyl transferase polypeptide (HPT1-1).

SEQ ID NO: 7 sets forth an *Arabidopsis thaliana* homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NO: 8 sets forth a partial *Cuphea pulcherrima* homogentisate prenyl transferase polypeptide.

SEQ ID NO: 9 sets forth a leek homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NO: 10 sets forth a wheat homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NO: 11 sets forth a *Cuphea pulcherrima* homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NOs: 12-15 represent domains from SEQ ID NOs: 1-8.

SEQ ID NOs: 16-26 set forth primer sequences.

SEQ ID NO: 27 sets forth a nucleic acid molecule encoding a *Nostoc punctiforme* homogentisate prenyl transferase polypeptide.

SEQ ID NO: 28 sets forth a nucleic acid molecule encoding an *Anabaena* homogentisate prenyl transferase polypeptide.

SEQ ID NO: 29 sets forth a nucleic acid molecule encoding a *Synechocystis* homogentisate prenyl transferase polypeptide.

SEQ ID NO: 30 sets forth a nucleic acid molecule encoding a *Zea mays* homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NO: 31 sets forth a nucleic acid molecule encoding a *Glycine max* homogentisate prenyl transferase polypeptide (HPT1-2).

SEQ ID NO: 32 sets forth a nucleic acid molecule encoding a *Glycine max* homogentisate prenyl transferase polypeptide (HPT1-1).

SEQ ID NO: 33 sets forth a nucleic acid molecule encoding an *Arabidopsis thaliana* homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NO: 34 sets forth a nucleic acid molecule encoding a *Cuphea pulcherrima* homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NO: 35 sets forth a nucleic acid molecule encoding a leek homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NO: 36 sets forth a nucleic acid molecule encoding a wheat homogentisate prenyl transferase polypeptide (HPT1).

SEQ ID NOs: 37-38 set forth primer sequences.

SEQ ID NOs: 39-42 set forth domains from SEQ ID NOs: 1-7 and 9-11.

SEQ ID NO: 43 sets forth a homogentisate prenyl transferase polypeptide from *Trichodesmium erythraeum*.

SEQ ID NO: 44 sets forth a homogentisate prenyl transferase polypeptide from *Chloroflexus aurantiacus*.

SEQ ID NO: 45 sets forth a putative sequence for an *Arabidopsis thaliana* homogentisate prenyl transferase polypeptide (HPT2).

SEQ ID NOs: 46-49 represent domains from SEQ ID NOs: 1-4, 6-7, 9-11, 57-58 and 91.

SEQ ID NOs: 50-56 set forth primer sequences.

SEQ ID NO: 57 sets forth an *Arabidopsis thaliana* homogentisate prenyl transferase polypeptide (HPT2).

SEQ ID NO: 58 sets forth an *Oryza sativa* homogentisate prenyl transferase polypeptide (HPT2).

SEQ ID NO: 59 sets forth a nucleic acid molecule encoding an *Arabidopsis thaliana* homogentisate prenyl transferase polypeptide (HPT2).

SEQ ID NO: 60 sets forth a nucleic acid molecule encoding an *Oryza sativa* homogentisate prenyl transferase polypeptide (HPT2).

SEQ ID NO: 61 sets forth a putative homogentisate prenyl transferase polypeptide from *Arabidopsis thaliana* (HPT2).

SEQ ID NO: 62 sets forth a putative homogentisate prenyl transferase polypeptide from *Arabidopsis thaliana* (HPT2).

SEQ ID NO: 63 sets forth an EST from *Arabidopsis thaliana*.

SEQ ID NO: 64 sets forth an EST from *Medicago truncatula*.

SEQ ID NO: 65 sets forth an EST from *Medicago truncatula* developing stem.

SEQ ID NO: 66 sets forth an EST from *Medicago truncatula* developing stem.

SEQ ID NO: 67 sets forth an EST from *Medicago truncatula* developing stem.

SEQ ID NO: 68 sets forth an EST from mixed potato tissues.

SEQ ID NO: 69 sets forth an EST from *Arabidopsis thaliana*, Columbia ecotype flower buds.

SEQ ID NO: 70 sets forth an EST from *Arabidopsis thaliana*.

SEQ ID NO: 71 sets forth an EST from *Medicago truncatula*.

SEQ ID NO: 72 sets forth an EST from *Glycine max*.

SEQ ID NOs: 73-83 and 84-87 set forth primer sequences.

SEQ ID NO: 88 sets forth a nucleic acid molecule encoding a homogentisate prenyl transferase polypeptide from cyanobacteria *Trichodesmium erythraeum*.

SEQ ID NO: 89 sets forth a nucleic acid molecule encoding a homogentisate prenyl transferase polypeptide from photobacteria *Chloroflexus aurantiacus*.

SEQ ID NO: 90 sets forth a *Glycine max* homogentisate prenyl transferase polypeptide (HPT2).

SEQ ID NO: 91 sets forth a nucleic acid molecule encoding a homogentisate prenyl transferase polypeptide from *Glycine max* (HPT2).

SEQ ID NOs: 92-95 represent domains from SEQ ID NOs: 1-4, 6-7, 9-11, 43-44, 57-58, and 90.

Note: cyanobacteria and photobbacteria have one HPT. Plants have both HPT1 and HPT2. In soy, there are two variations of HPT1, HPT1-1 and HPT1-2, as well as HPT2.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a schematic diagram of the tocopherol biosynthetic pathway.

Figures 2a-2c depicts a sequence alignment for several homogentisate prenyl transferase polypeptides SEQ ID NOs: 1-8).

Figures 3a-3c depicts a sequence alignment for several homogentisate prenyl transferase polypeptides (SEQ ID NOs: 1-7, and 9-11).

Figure 4 provides a schematic of the expression construct pCGN10800.

Figure 5 provides a schematic of the expression construct pCGN10801.

Figure 6 provides a schematic of the expression construct pCGN10803.

Figure 7 provides a schematic of the expression construct pCGN10822.

Figure 8 provides bar graphs of HPLC data obtained from seed extracts of transgenic *Arabidopsis* containing pCGN10822, which provides of the expression of the ATPT2 sequence (SEQ ID NO: 33), in the sense orientation, from the napin promoter. Provided are graphs for α, γ, and δ-tocopherols, as well as total tocopherol for 22 transformed lines, as well as a nontransformed (wild-type) control.

Figure 9 provides a bar graph of HPLC analysis of seed extracts from *Arabidopsis* plants transformed with a pCGN10803 (lines 1387 through 1624, enhanced 35S-ATPT2, in the antisense orientation), a nontransformed (wt) control, and an empty vector transformed control.

Figure 10 provides a schematic of the expression construct pMON36581.

Figure 11 provides a schematic of the expression construct pMON69933.

Figure 12 provides a schematic of the expression construct pMON69924.

Figure 13 provides a schematic of the expression construct pMON69943.

Figure 14 provides a bar graph of total tocopherol levels in recombinant soy lines.

Figure 15 depicts pMON 69960.

Figure 16 depicts pMON 36525.

Figure 17 depicts pMON 69963.

Figure 18 depicts pMON 69965.

Figure 19 depicts pMON 10098.

Figure 20 depicts pMON 69964.

Figure 21 depicts pMON 69966.

Figure 22 depicts results of seed total tocopherol analysis.

Figure 23 depicts results of seed total tocopherol analysis.

Figure 24 depicts the alignments of SEQ ID NOs: 1-4, 6-7, 9-11, 57, and 90.

Figure 25 depicts motifs V through VIII, SEQ ID NOs: 46-49.

Figure 26 depicts a sequence tree derived from a multiple alignment shown from SEQ ID NOs: 1-7, 9-11, 43, 44, 57-58, and 90.

Figure 27 depicts pMON81028.

Figure 28 depicts pMON81023.

Figure 29 depicts pMON36596.

Figure 30 depicts pET30a(+) vector.

Figure 31 depicts pMON69993.

Figure 32 depicts pMON69992.

Figures 33a-33c depicts a sequence alignment for several homogentisate prenyl transferase polypeptide SEQ ID NOs: 1-4, 6-7, 9-11, 43-44, 57-58, and 90.

Figure 34 depicts motifs IX through XII, SEQ ID NOs: 92-95.

Figure 35 depicts motifs I-IV, SEQ ID NOs: 39-42.

Figure 36 depicts motifs A-D.

DETAILED DESCRIPTION

The present invention provides a number of agents, for example, nucleic acid molecules and polypeptides associated with the synthesis of tocopherol, and provides uses of such agents.

AGENTS

The agents of the present invention will preferably be "biologically active" with respect to either a structural attribute, such as the capacity of a nucleic acid to hybridize to another nucleic acid molecule, or the ability of a protein to be bound by an antibody (or to compete with another molecule for such binding). Alternatively, such an attribute may be catalytic and thus involve the capacity of the agent to mediate a chemical reaction or response. The agents will preferably be "substantially purified". The term "substantially purified", as used herein, refers to a molecule separated from substantially all other molecules normally associated with it in its native environmental conditions. More preferably a substantially purified molecule is the predominant species present in a preparation. A substantially purified molecule may be greater than about 60% free, preferably about 75% free, more preferably about 90% free, and most preferably about 95% free from the other molecules (exclusive of solvent) present in the natural mixture. The term "substantially purified" is not intended to encompass molecules present in their native environmental conditions.

The agents of the present invention may also be recombinant. As used herein, the term recombinant means any agent (*e.g.*, DNA, peptide *etc.*), that is, or results, however indirectly, from human manipulation of a nucleic acid molecule.

It is understood that the agents of the present invention may be labeled with reagents that facilitate detection of the agent (*e.g.*, fluorescent labels, Prober *et al., Science*, 238:336-340 (1987); Albarella *et al.*, EP 144 914; chemical labels, Sheldon *et al.*, U.S. Patent 4,582,789; Albarella *et al.*, U.S. Patent 4,563,417; modified bases, Miyoshi *et al.*, EP 119 448).

NUCLEIC ACID MOLECULES

Agents of the present invention include nucleic acid molecules. In a preferred aspect of the present invention the nucleic acid molecule comprises a nucleic acid sequence that encodes a homogentisate prenyl transferase. As used herein, a homogentisate prenyl transferase is any plant protein that is capable of specifically catalyzing the formation of 2-methyl-6-phytylbenzoquinol (2-methyl-6-geranylgeranylbenzoquinol) from phytyl-DP (GGDP) and homogentisate.

An example of a more preferred homogentisate prenyl transferase is a polypeptide with the amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11,

43-44, 55, 58, and 90. In a more preferred embodiment, the homogentisate prenyl transferase is encoded by any nucleic acid molecule encoding an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 55, 58, and 90.

In another preferred aspect of the present invention the nucleic acid molecule of the present invention comprises a nucleic acid sequence encoding a polypeptide selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 55, 58, and 90, and complements thereof and fragments of either.

In another preferred aspect of the present invention the nucleic acid molecule of the present invention comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 31, 34-36, 59-60, and 91.

In another embodiment, the present invention includes nucleic acid molecules encoding polypeptides having a region of conserved amino acid sequence shown in any of Figures Figures 2a-2c, 3a-3c, 24a-24b, 25a-25b, 33a-33c, 34a-34b, 35a-b and 36, and complements of those nucleic acid molecules. In a preferred embodiment, the present invention includes nucleic acid molecules encoding polypeptides comprising a sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95, and complements of those nucleic acid molecules. The present invention includes and provides said nucleic acid molecule wherein the polypeptide further comprises more than one amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In a further preferred embodiment the present invention includes nucleic acid molecules encoding polypeptides comprising two or more, three or more, or four sequences selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95, and complements of those nucleic acid molecules. In another embodiment, the present invention includes nucleic acid molecules encoding polypeptides having homogentisate prenyl transferase activity and a region of conserved amino acid sequence shown in any of Figures 2a-2c, 3a-3c, 24a-24b, 25a-25b, 33a-33c, 34a-34b, 35a-35b and 36, and complements of those nucleic acid molecules. In a preferred embodiment, the present invention includes nucleic acid molecules encoding polypeptides having homogentisate prenyl transferase activity and comprising a sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95 and complements of those nucleic acid molecules. The present invention includes and provides said nucleic acid molecule wherein the polypeptide further comprises more than one

amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In a further preferred embodiment the present invention includes nucleic acid molecules encoding polypeptides having homogentisate prenyl transferase activity and comprising two or more, three or more, or four sequences selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95, and complements of those nucleic acid molecules. In another embodiment, the present invention includes nucleic acid molecules, excluding nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato, encoding polypeptides having a region of conserved amino acid sequence shown in any of Figures 2a-2c, 3a-3c, 24a-24b, 25a-25b, 33a-33c, 34a-34b, 35a-35b and 36, and complements of those nucleic acid molecules. In a preferred embodiment, the present invention includes nucleic acid molecules, excluding nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato, encoding polypeptides comprising a sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.and complements of those nucleic acid molecules. The present invention includes and provides said nucleic acid molecule wherein the polypeptide further comprises more than one amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In a further preferred embodiment the present invention includes nucleic acid molecules, excluding nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato, encoding polypeptides comprising two or more, three or more, or four sequences selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In another embodiment, the present invention includes nucleic acid molecules, excluding nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, Sulfolobus, Aeropyum, sorghum, or tomato, encoding polypeptides having homogentisate prenyl transferase activity and a region of conserved amino acid sequence shown in any of Figures 2a-2c, 3a-3c, 24a-24b, 25a-25b,

33a-33c, 34a-34b, 35a-35b and 36 and complements of those nucleic acid molecules. In a preferred embodiment, the present invention includes nucleic acid molecules, excluding nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato, encoding polypeptides having homogentisate prenyl transferase activity and comprising a sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95. The present invention includes and provides said nucleic acid molecule wherein the polypeptide further comprises more than one amino acid sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In a further preferred embodiment the present invention includes nucleic acid molecules, excluding nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato, encoding polypeptides having homogentisate prenyl transferase activity and comprising two or more, three or more, or four sequences selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In one embodiment of a method of the present invention, any of the nucleic acid sequences or polypeptide sequences, or fragments of either, of the present invention can be used to search for related sequences. In a preferred embodiment, a member selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90 is used to search for related sequences. In a preferred embodiment, a member selected from the group consisting of SEQ ID NOs: 31, 34-36, 59-60, 88-89, and 91 is used to search for related sequences. In another embodiment, any of the motifs or regions of conserved sequence shown in Figures 2a-2c, 3a-3c, 24a-24b, 25a-25b, 33a-33c, 34a-34b, 35a-35b and 36 are used to search for related amino acid sequences. In a preferred embodiment, a member selected from the group consisting of SEQ ID NOs: 39-42 and 46-49 is used to search for related sequences. In one embodiment, one or more of SEQ ID NOs: 39-42, 46-49, and 92-95 is used to search for related sequences. As used herein, "search for related sequences" means any method of determining relatedness between two sequences, including, but not limited to, searches that compare sequence homology: for example, a PBLAST search of a database for relatedness to a single amino acid sequence. Other searches may be conducted using profile based methods,

such as the HMM (Hidden Markov model) META-MEME, PSI-BLAST.

The present invention includes and provides for

homogentisate prenyl transferases discovered using one or more of the alignments of Figures 2a-2c, 3a-3c, 24a-24b, 25a-25b, 33a-33c, 34a-34b, 35a-35b and 36.

As used herein, a nucleic acid molecule is said to be "derived from" a particular organism, species, ecotype, etc., when the sequence of the nucleic acid molecule originated from that organism, species, ecotype, etc. "Derived from" therefore includes copies of nucleic acid molecules derived through, for example, PCR, as well as synthetically generated nucleic acid molecules having the same nucleic acid sequence as the original organism, species, ecotype, etc. Likewise, a polypeptide is said to be "derived from" a nucleic acid molecule when that nucleic acid molecule is used to code for the polypeptide, whether the polypeptide is enzymatically generated from the nucleic acid molecule or synthesized based on the sequence information inherent in the nucleic acid molecule.

The present invention includes the use of the above-described conserved sequences and fragments thereof in transgenic plants, other organisms, and for other uses, including, without limitation, as described below.

In another preferred aspect of the present invention a nucleic acid molecule comprises nucleotide sequences encoding a plastid transit peptide operably fused to a nucleic acid molecule that encodes a protein or fragment of the present invention.

In another preferred embodiment of the present invention, the nucleic acid molecules of the present invention encode mutant tocopherol homogentisate prenyl transferase enzymes. As used herein, a mutant enzyme is any enzyme that contains an amino acid that is different from the amino acid in the same position of a wild type enzyme of the same type.

It is understood that in a further aspect of nucleic acid sequences of the present invention, the nucleic acids can encode a protein that differs from any of the proteins in that one or more amino acids have been deleted, substituted or added without altering the function. For example, it is understood that codons capable of coding for such conservative amino acid substitutions are known in the art.

In one aspect of the present invention the nucleic acids of the present invention are said to be introduced nucleic acid molecules. A nucleic acid molecule is said to be "introduced" if it is inserted into a cell or organism as a result of human manipulation, no

matter how indirect. Examples of introduced nucleic acid molecules include, without limitation, nucleic acids that have been introduced into cells via transformation, transfection, injection, and projection, and those that have been introduced into an organism via conjugation, endocytosis, phagocytosis, etc.

One subset of the nucleic acid molecules of the present invention is fragment nucleic acids molecules. Fragment nucleic acid molecules may consist of significant portion(s) of, or indeed most of, the nucleic acid molecules of the present invention, such as those specifically disclosed. Alternatively, the fragments may comprise smaller oligonucleotides (having from about 15 to about 400 nucleotide residues and more preferably, about 15 to about 30 nucleotide residues, or about 50 to about 100 nucleotide residues, or about 100 to about 200 nucleotide residues, or about 200 to about 400 nucleotide residues, or about 275 to about 350 nucleotide residues).

A fragment of one or more of the nucleic acid molecules of the present invention may be a probe and specifically a PCR probe. A PCR probe is a nucleic acid molecule capable of initiating a polymerase activity while in a double-stranded structure with another nucleic acid. Various methods for determining the structure of PCR probes and PCR techniques exist in the art. Computer generated searches using programs such as Primer3, STSPipeline, or GeneUp (Pesole *et al., BioTechniques*, 25:112-123 (1998), for example, can be used to identify potential PCR primers.

Nucleic acid molecules or fragments thereof of the present invention are capable of specifically hybridizing to other nucleic acid molecules under certain circumstances. Nucleic acid molecules of the present invention include those that specifically hybridize to those nucleic acid molecules disclosed herein, such as those encoding any of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof. Nucleic acid molecules of the present invention include those that specifically hybridize to a nucleic acid molecules comprising a member selected from the group consisting of SEQ ID NOs: 31, 34-36, 59-60, and 91, and complements thereof.

As used herein, two nucleic acid molecules are said to be capable of specifically hybridizing to one another if the two molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure.

A nucleic acid molecule is said to be the "complement" of another nucleic acid molecule if they exhibit complete complementarity. As used herein, molecules are said to exhibit "complete complementarity" when every nucleotide of one of the molecules is complementary to a nucleotide of the other. Two molecules are said to be "minimally complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under at least conventional "low-stringency" conditions. Similarly, the molecules are said to be "complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under conventional "high-stringency" conditions. Conventional stringency conditions are described by Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual, 2nd Ed., Cold Spring Harbor Press*, Cold Spring Harbor, NY (1989), and by Haymes *et al., Nucleic Acid Hybridization, A Practical Approach, IRL Press*, Washington, DC (1985). Departures from complete complementarity are therefore permissible, as long as such departures do not completely preclude the capacity of the molecules to form a double-stranded structure. Thus, in order for a nucleic acid molecule to serve as a primer or probe it need only be sufficiently complementary in sequence to be able to form a stable double-stranded structure under the particular solvent and salt concentrations employed.

Appropriate stringency conditions which promote DNA hybridization are, for example, 6.0 X sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 X SSC at 20-25°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, NY (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 X SSC at 50°C to a high stringency of about 0.2 X SSC at 65°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Both temperature and salt may be varied, or either the temperature or the salt concentration may be held constant while the other variable is changed.

In a preferred embodiment, a nucleic acid of the present invention will specifically hybridize to one or more of the nucleic acid molecules described herein and complements thereof, such as those encoding any of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, under moderately stringent conditions, for example at about 2.0 X SSC and about 65°C.

In a particularly preferred embodiment, a nucleic acid of the present invention will include those nucleic acid molecules that specifically hybridize to one or more nucleic acid molecules encoding any of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof, under high stringency conditions such as 0.2 X SSC and about 65°C.

In one aspect of the present invention, the nucleic acid molecules of the present invention have one or more nucleic acid sequences encoding SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, or complements thereof. In another aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between about 100% and about 90% sequence identity with one or more of the nucleic acid sequences encoding SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof, and fragments of either. In a further aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between about 100% and about 95% sequence identity with one or more of the nucleic acid sequences encoding SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof, and fragments of either. In a more preferred aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between about 100% and about 98% sequence identity with one or more of the nucleic acid sequences encoding SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof, and fragments of either. In an even more preferred aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between about 100% and about 99% sequence identity with one or more of the sequences encoding SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof, and fragments of either.

In a preferred embodiment the percent identity calculations are performed using BLASTN or BLASTP (default, parameters, version 2.0.8, Altschul *et al., Nucleic Acids Res.*, 25:3389-3402 (1997)).

A nucleic acid molecule of the present invention can also encode a homolog polypeptide. As used herein, a homolog polypeptide molecule or fragment thereof is a counterpart protein molecule or fragment thereof in a second species (*e.g.*, corn rubisco small subunit is a homolog of *Arabidopsis* rubisco small subunit). A homolog can also be generated by molecular evolution or DNA shuffling techniques, so that the molecule retains at least one functional or structure characteristic of the original polypeptide (*see*, for example, U.S. Patent 5,811,238).

In another embodiment, the homolog is selected from the group consisting of alfalfa, *Arabidopsis*, barley, *Brassica campestris, Brassica napus,* oilseed rape, broccoli, cabbage, canola, citrus, cotton, garlic, oat, Allium, flax, an ornamental plant, peanut, pepper, potato, rapeseed, rice, rye, sorghum, strawberry, sugarcane, sugarbeet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, soybean, corn, *Phaseolus,* crambe, mustard, castor bean, sesame, cottonseed, linseed, safflower, and oil palm. More particularly, preferred homologs are selected from canola, corn, *Brassica campestris, Brassica napus,* oilseed rape, soybean, crambe, mustard, castor bean, peanut, sesame, cottonseed, linseed, rapeseed, safflower, oil palm, flax, and sunflower. In an even more preferred embodiment, the homolog is selected from the group consisting of canola, rapeseed, corn, *Brassica campestris, Brassica napus,* oilseed rape, soybean, sunflower, safflower, oil palms, and peanut. In a particularly preferred embodiment, the homolog is soybean. In a particularly preferred embodiment, the homolog is canola. In a particularly preferred embodiment, the homolog is oilseed rape.

In a preferred embodiment, nucleic acid molecules encoding SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof, and fragments of either; or more preferably encoding SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof, can be utilized to obtain such homologs.

In another further aspect of the present invention, nucleic acid molecules of the present invention can comprise sequences that differ from those encoding a polypeptide or fragment thereof due to the fact that a polypeptide can have one or more conservative amino acid changes, and nucleic acid sequences coding for the polypeptide can therefore have sequence differences. It is understood that codons capable of coding for such conservative amino acid substitutions are known in the art.

It is well known in the art that one or more amino acids in a native sequence can be substituted with other amino acid(s), the charge and polarity of which are similar to that of the native amino acid, *i.e.*, a conservative amino acid substitution. Conservative substitutes for an amino acid within the native polypeptide sequence can be selected from other members of the class to which the amino acid belongs. Amino acids can be divided into the following four groups: (1) acidic amino acids; (2) basic amino acids; (3) neutral polar amino acids; and (4) neutral nonpolar amino acids. Representative amino acids within these various groups include, but are not limited to, (1) acidic (negatively charged) amino acids such as aspartic

acid and glutamic acid; (2) basic (positively charged) amino acids such as arginine, histidine, and lysine; (3) neutral polar amino acids such as glycine, serine, threonine, cysteine, cystine, tyrosine, asparagine, and glutamine; and (4) neutral nonpolar (hydrophobic) amino acids such as alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine.

Conservative amino acid substitution within the native polypeptide sequence can be made by replacing one amino acid from within one of these groups with another amino acid from within the same group. In a preferred aspect, biologically functional equivalents of the proteins or fragments thereof of the present invention can have ten or fewer conservative amino acid changes, more preferably seven or fewer conservative amino acid changes, and most preferably five or fewer conservative amino acid changes. The encoding nucleotide sequence will thus have corresponding base substitutions, permitting it to encode biologically functional equivalent forms of the polypeptides of the present invention.

It is understood that certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Because it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence and, of course, its underlying DNA coding sequence and, nevertheless, a protein with like properties can still be obtained. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the proteins or fragments of the present invention, or corresponding DNA sequences that encode said peptides, without appreciable loss of their biological utility or activity. It is understood that codons capable of coding for such amino acid changes are known in the art.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte and Doolittle, *J. Mol. Biol.*, 157:105-132 (1982)). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant polypeptide, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, *J. Mol. Biol.*, 157:105-132

(1982)); these are isoleucine (+4.5), valine (+4.2), leucine (+3.8), phenylalanine (+2.8), cysteine/cystine (+2.5), methionine (+1.9), alanine (+1.8), glycine (-0.4), threonine (-0.7), serine (-0.8), tryptophan (-0.9), tyrosine (-1.3), proline (-1.6), histidine (-3.2), glutamate (-3.5), glutamine (-3.5), aspartate (-3.5), asparagine (-3.5), lysine (-3.9), and arginine (-4.5).

In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent 4,554,101 states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0), lysine (+3.0), aspartate (+3.0±1), glutamate (+3.0±1), serine (+0.3), asparagine (+0.2), glutamine (+0.2), glycine (0), threonine (-0.4), proline (-0.5±1), alanine (-0.5), histidine (-0.5), cysteine (-1.0), methionine (-1.3), valine (-1.5), leucine (-1.8), isoleucine (-1.8), tyrosine (-2.3), phenylalanine (-2.5), and tryptophan (-3.4).

In making such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

In a further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those for which a specific sequence is provided herein because one or more codons has been replaced with a codon that encodes a conservative substitution of the amino acid originally encoded.

Agents of the present invention include nucleic acid molecules that encode at least about a contiguous 10 amino acid region of a polypeptide of the present invention, more preferably at least about a contiguous 25, 40, 50, 100, or 125 amino acid region of a polypeptide of the present invention.

In a preferred embodiment, any of the nucleic acid molecules of the present invention can be operably linked to a promoter region that functions in a plant cell to cause the production of an mRNA molecule, where the nucleic acid molecule that is linked to the

promoter is heterologous with respect to that promoter. As used herein, "heterologous" means not naturally occurring together.

The nature of the coding sequences of non-plant genes can distinguish them from plant genes as well as many other heterologous genes expressed in plants. For example, the average A+T content of bacteria can be higher than that for plants. The A+T content of the genomes (and thus the genes) of any organism are features of that organism and reflect its evolutionary history. While within any one organism genes have similar A+T content, the A+T content can vary tremendously from organism to organism. For example, some Bacillus species have among the most A+T rich genomes while some Steptomyces species are among the least A+T rich genomes (about 30 to 35% A+T).

Due to the degeneracy of the genetic code and the limited number of codon choices for any amino acid, most of the "excess" A+T of the structural coding sequences of some Bacillus species, for example, are found in the third position of the codons. That is, genes of some Bacillus species have A or T as the third nucleotide in many codons. Thus A+T content in part can determine codon usage bias. In addition, it is clear that genes evolve for maximum function in the organism in which they evolve. This means that particular nucleotide sequences found in a gene from one organism, where they may play no role except to code for a particular stretch of amino acids, have the potential to be recognized as gene control elements in another organism (such as transcriptional promoters or terminators, polyA addition sites, intron splice sites, or specific mRNA degradation signals). It is perhaps surprising that such misread signals are not a more common feature of heterologous gene expression, but this can be explained in part by the relatively homogeneous A+T content (about 50%) of many organisms. This A+T content plus the nature of the genetic code put clear constraints on the likelihood of occurrence of any particular oligonucleotide sequence. Thus, a gene from *E. coli* with a 50% A+T content is much less likely to contain any particular A+T rich segment than a gene from *B. thuringiensis*. The same can be true between genes in a bacterium and genes in a plant, for example.

Any of the nucleic acid molecules of the present invention can be altered via any methods known in the art in order to make the codons within the nucleic acid molecule more appropriate for the organism in which the nucleic acid molecule is located. That is, the present invention includes the modification of any of the nucleic acid molecules disclosed herein to improve codon usage in a host organism.

It is preferred that regions comprising many consecutive A+T bases or G+C bases are disrupted since these regions are predicted to have a higher likelihood to form hairpin structure due to self-complementarity. Therefore, insertion of heterogeneous base pairs would reduce the likelihood of self-complementary secondary structure formation which are known to inhibit transcription and/or translation in some organisms. In most cases, the adverse effects may be minimized by using sequences which do not contain more than five consecutive A+T or G+C.

PROTEIN AND PEPTIDE MOLECULES

A class of agents includes one or more of the polypeptide molecules encoded by a nucleic acid agent of the present invention. A particular preferred class of proteins is that having an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and fragments thereof.

In another embodiment, the present invention includes polypeptides having a region of conserved amino acid sequence shown in any of Figures 2a-2c, 3a-3c, 24a-24b, 25a-25b, 33a-33c, 34a-34b, 35a-35b and 36. In an embodiment, the present invention includes polypeptides comprising a sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95. The present invention includes and provides said substantially purified polypeptide wherein more than one amino acid sequence is selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95. In a further preferred embodiment the present invention includes polypeptides comprising two or more, three or more, or four sequences selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In another embodiment, the present invention includes polypeptides having homogentisate prenyl transferase activity and a region of conserved amino acid sequence shown in any of Figures 2a-2c, 3a-3c, 25a-25c, 33a-33c, 34a-34b, 35a-35b and 36. In an embodiment, the present invention includes polypeptides having homogentisate prenyl transferase activity and comprising a sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95. The present invention includes and provides said substantially purified polypeptide wherein more than one amino acid sequence is selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In a further preferred embodiment the present invention includes polypeptides having homogentisate prenyl transferase activity and comprising two or more, three or more, or four sequences selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In another embodiment, the present invention includes polypeptides having a region of conserved amino acid sequence shown in any of Figures 2a-2c, 3a-3c, 25a-25c, 33a-33c, 34a-34b, 35a-35b or 36, excluding polypeptides derived from nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, Sulfolobus, Aeropyum, sorghum, or tomato. In a preferred embodiment, the present invention includes polypeptides comprising a sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95 excluding polypeptides derived from nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato. The present invention includes and provides said substantially purified polypeptide wherein more than onethe amino acid sequence is selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In a further preferred embodiment the present invention includes polypeptides comprising two or more, three or more, or four sequences selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95, excluding polypeptides derived from nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato.

In another embodiment, the present invention includes polypeptides having homogentisate prenyl transferase activity and a region of conserved amino acid sequence shown in any of Figures 2a-2c, 3a-3c, 25a-25c, 33a-33c, 34a-34b, 35a-35b or 36, excluding polypeptides derived from nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato. In a preferred embodiment, the present invention includes polypeptides having homogentisate prenyl transferase activity and comprising a sequence selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95, excluding polypeptides derived from nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato. The present invention includes and

provides said substantially purified polypeptide wherein more than one amino acid sequence is selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95.

In a further preferred embodiment the present invention includes polypeptides having homogentisate prenyl transferase activity and comprising two or more, three or more, or four sequences selected from the group consisting of SEQ ID NOs: 39-42, 46-49, and 92-95, excluding polypeptides derived from nucleic acid molecules derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa, Trichodesmium erythraeum, Chloroflexus aurantiacus,* wheat, leek, canola, cotton, or tomato.

Polypeptide agents may have C-terminal or N-terminal amino acid sequence extensions. One class of N-terminal extensions employed in a preferred embodiment are plastid transit peptides. When employed, plastid transit peptides can be operatively linked to the N-terminal sequence, thereby permitting the localization of the agent polypeptides to plastids. In an embodiment of the present invention, any suitable plastid targeting sequence can be used. Where suitable, a plastid targeting sequence can be substituted for a native plastid targetting sequence, for example, for the CTP occurring natively in the tocopherol homogentisate prenyl transferase protein. In a further embodiment, a plastid targeting sequence that is heterologous to any homogentisate prenyl transferase protein or fragment described herein can be used. In a further embodiment, any suitable, modified plastid targetting sequence can be used. In another embodiment, the plastid targeting sequence is a CTP1 sequence (*see* WO 00/61771).

In a preferred aspect a protein of the present invention is targeted to a plastid using either a native transit peptide sequence or a heterologous transit peptide sequence. In the case of nucleic acid sequences corresponding to nucleic acid sequences of non-higher plant organisms such as cynobacteria, such nucleic acid sequences can be modified to attach the coding sequence of the protein to a nucleic acid sequence of a plastid targeting peptide.

As used herein, the terms "protein", "peptide molecule", or "polypeptide" include any molecule that comprises five or more amino acids. It is well known in the art that protein, peptide, or polypeptide molecules may undergo modification, including post-translational modifications, such as, but not limited to, disulfide bond formation, glycosylation, phosphorylation, or oligomerization. Thus, as used herein, the terms "protein", "peptide molecule", or "polypeptide" include any protein that is modified by any biological or non-

biological process. The terms "amino acid" and "amino acids" refer to all naturally occurring L-amino acids. This definition is meant to include norleucine, norvaline, ornithine, homocysteine, and homoserine.

One or more of the protein or fragments thereof, peptide molecules, or polypeptide molecules may be produced via chemical synthesis, or more preferably, by expression in a suitable bacterial or eukaryotic host. Suitable methods for expression are described by Sambrook *et al.*, In: *Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press*, Cold Spring Harbor, NY (1989) or similar texts.

A "protein fragment" is a peptide or polypeptide molecule whose amino acid sequence comprises a subset of the amino acid sequence of that protein. A protein or fragment thereof that comprises one or more additional peptide regions not derived from that protein is a "fusion" protein. Such molecules may be derivatized to contain carbohydrate or other moieties (such as keyhole limpet hemocyanin). Fusion protein or peptide molecules of the present invention are preferably produced via recombinant means.

Another class of agents comprises protein, peptide molecules, or polypeptide molecules, or fragments or fusions thereof comprising SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and fragments thereof in which conservative, non-essential, or non-relevant amino acid residues have been added, replaced, or deleted. Computerized means for designing modifications in protein structure are known in the art (Dahiyat and Mayo, *Science*, 278:82-87 (1997)).

A protein, peptide, or polypeptide of the present invention can also be a homolog protein, peptide, or polypeptide. As used herein, a homolog protein, peptide, or polypeptide or fragment thereof is a counterpart protein, peptide, or polypeptide or fragment thereof in a second species. A homolog can also be generated by molecular evolution or DNA shuffling techniques, so that the molecule retains at least one functional or structure characteristic of the original (*see*, for example, U.S. Patent 5,811,238).

In another embodiment, the homolog is selected from the group consisting of alfalfa, *Arabidopsis*, barley, broccoli, cabbage, canola, citrus, cotton, garlic, oat, Allium, flax, an ornamental plant, peanut, pepper, potato, rapeseed, rice, rye, sorghum, strawberry, sugarcane, sugarbeet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, soybean, corn, and *Phaseolus*. More particularly, preferred homologs are selected from canola, rapeseed, corn, *Brassica campestris, Brassica napus,*

oilseed rape, soybean, crambe, mustard, castor bean, peanut, sesame, cottonseed, linseed, safflower, oil palm, flax, and sunflower. In an even more preferred embodiment, the homolog is selected from the group consisting of canola, rapeseed, corn, *Brassica campestris, Brassica napus,* oilseed rape, soybean, sunflower, safflower, oil palms, and peanut. In a preferred embodiment, the homolog is soybean. In a preferred embodiment, the homolog is canola. In a preferred embodiment, the homolog is oilseed rape.

In a preferred embodiment, the nucleic acid molecules of the present invention or complements and fragments of either can be utilized to obtain such homologs.

Agents of the present invention include proteins and fragments thereof comprising at least about a contiguous 10 amino acid region preferably comprising at least about a contiguous 20 amino acid region, even more preferably comprising at least about a contiguous 25, 35, 50, 75, or 100 amino acid region of a protein of the present invention. In another preferred embodiment, the proteins of the present invention include between about 10 and about 25 contiguous amino acid region, more preferably between about 20 and about 50 contiguous amino acid region, and even more preferably between about 40 and about 80 contiguous amino acid region.

PLANT CONSTRUCTS AND PLANT TRANSFORMANTS

One or more of the nucleic acid molecules of the present invention may be used in plant transformation or transfection. Exogenous genetic material may be transferred into a plant cell and the plant cell regenerated into a whole, fertile, or sterile plant. Exogenous genetic material is any genetic material, whether naturally occurring or otherwise, from any source that is capable of being inserted into any organism.

In a preferred aspect of the present invention the exogenous genetic material comprises a nucleic acid sequence of the present invention, more preferably one that encodes homogentisate prenyl transferase. In another preferred aspect of the present invention the exogenous genetic material of the present invention comprises a nucleic acid sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof and fragments of either. In a further aspect of the present invention the exogenous genetic material comprises a nucleic acid sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and fragments of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90.

In an embodiment of the present invention, exogenous genetic material encoding a homogentisate prenyl transferase enzyme or fragment thereof is introduced into a plant with one or more additional genes. In one embodiment, preferred combinations of genes include a nucleic acid molecule of the present invention and one or more of the following genes: *tyrA* (*e.g.*, WO 02/089561 and Xia *et al.*, *J. Gen. Microbiol.*, 138:1309-1316 (1992)), *tocopherol cyclase* (*e.g.*, WO 01/79472), *prephenate dehydrogenase*, *dxs* (*e.g.* Lois *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 95(5):2105-2110 (1998)), *dxr* (*e.g.*, U.S. Pub. 2002/0108814A and Takahashi *et al., Proc. Natl. Acad. Sci. (U.S.A.)*, 95 (17), 9879-9884 (1998)), *GGPPS* (*e.g.*, Bartley and Scolnik, *Plant Physiol.*, 104:1469-1470 (1994)), *HPPD* (*e.g.*, Norris *et al.*, *Plant Physiol.*, 117:1317-1323 (1998)), *GMT* (*e.g.*, U.S. Application 10/219,810, filed August 16, 2002), *tMT2* (*e.g.*, U.S. Application 10/279,029, filed October 24, 2002), *AANT1* (*e.g.*, WO 02/090506), *IDI* (E.C.:5.3.3.2; Blanc *et al.*, In: *Plant Gene Register*, PRG96-036; and Sato *et al.*, *DNA Res.*, 4:215-230 (1997)), *GGH* (Graβes *et al.*, *Planta.* 213-620 (2001)), or a plant ortholog and an antisense construct for homogentisic acid dioxygenase (Kridl *et al.*, *Seed Sci. Res.*, 1:209:219 (1991); Keegstra, *Cell*, 56(2):247-53 (1989); Nawrath, *et al., Proc. Natl. Acad. Sci. (U.S.A.)*, 91:12760-12764 (1994); Cyanobase;
Smith *et al., Plant J.*, 11:83-92 (1997); WO 00/32757; ExPASy Molecular Biology Server, http://us.expasy.org/enzyme; *MT1* WO 00/10380; *gcpE*, WO 02/12478; Saint Guily *et al.*, *Plant Physiol.*, 100(2):1069-1071 (1992); Sato *et al.*, *J. DNA Res.*, 7(1):31-63 (2000)). In such combinations, in some crop plants, e.g., canola, a preferred promoter is a napin promoter and a preferred plastid targeting sequence is a CTP1 sequence. It is preferred that gene products are targeted to the plastid.

In a preferred combination a nucleic acid molecule encoding a homogentisate prenyl transferase polypeptide and a nucleic acid molecule encoding any of the following enzymes: *tyrA, prephenate dehydrogenase, tocopherol cyclase, dxs, dxr, GGPPS, HPPD, tMT2, MT1, GCPE, AANT1, IDI, GGH, GMT,* or a plant ortholog and an antisense construct for homogentisic acid dioxygenase are introduced into a plant.

For any of the above combinations, a nucleic acid molecule encoding a homogentisate prenyl transferase polypeptide encodes a polypeptide comprising a sequence selected from the group consisting of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90. In another preferred embodiment, a nucleic acid molecule encoding a homogentisate prenyl transferase polypeptide encodes a polypeptide comprising one or more of SEQ ID NOs: 39-42, 46-49,

and 92-95. In a preferred embodiment, the homogentisate prenyl transferase polypeptide does not have an amino acid sequence that is derived from a nucleic acid derived from *Nostoc punctiforme, Anabaena, Synechocystis, Zea mays, Glycine max, Arabidopsis thaliana, Oryza sativa*, wheat, leek, canola, cotton, or tomato.

Such genetic material may be transferred into either monocotyledons or dicotyledons including, but not limited to canola, corn, soybean, *Arabidopsis* phaseolus, peanut, alfalfa, wheat, rice, oat, sorghum, rapeseed, rye, tritordeum, millet, fescue, perennial ryegrass, sugarcane, cranberry, papaya, banana, safflower, oil palms, flax, muskmelon, apple, cucumber, dendrobium, gladiolus, chrysanthemum, liliacea, cotton, eucalyptus, sunflower, *Brassica campestris, Brassica napus,* oilseed rape, turfgrass, sugarbeet, coffee and dioscorea (Christou, In: *Particle Bombardment for Genetic Engineering of Plants,* Biotechnology Intelligence Unit. Academic Press, San Diego, CA (1996)), with canola, corn, *Brassica campestris, Brassica napus,* oilseed rape, rapeseed, soybean, crambe, mustard, castor bean, peanut, sesame, cottonseed, linseed, safflower, oil palm, flax, and sunflower preferred, and canola, rapeseed, corn, *Brassica campestris, Brassica napus,* oilseed rape, soybean, sunflower, safflower, oil palms, and peanut preferred. In a more preferred embodiment, the genetic material is transferred into canola. In another more preferred embodiment, the genetic material is transferred into oilseed rape. In another particularly preferred embodiment, the genetic material is transferred into soybean.

Transfer of a nucleic acid molecule that encodes a protein can result in expression or overexpression of that polypeptide in a transformed cell or transgenic plant. One or more of the proteins or fragments thereof encoded by nucleic acid molecules of the present invention may be overexpressed in a transformed cell or transformed plant. Such expression or overexpression may be the result of transient or stable transfer of the exogenous genetic material.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of tocopherols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of α-tocopherols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of γ-tocopherols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of δ-tocopherols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of β-tocopherols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of tocotrienols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of α-tocotrienols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of γ-tocotrienols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of δ-tocotrienols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of β-tocotrienols.

In a preferred embodiment, expression or overexpression of a polypeptide of the present invention in a plant provides in that plant, relative to an untransformed plant with a similar genetic background, an increased level of plastoquinols.

In any of the embodiments described herein, an increase in γ-tocopherol, α-tocopherol, or both can lead to a decrease in the relative proportion of β-tocopherol, δ-tocopherol, or both. Similarly, an increase in γ-tocotienol, α-tocotrienol, or both can lead to a decrease in the relative proportion of β-tocotrienol, δ-tocotrienol, or both.

In another embodiment, expression overexpression of a polypeptide of the present invention in a plant provides in that plant, or a tissue of that plant, relative to an untransformed plant or plant tissue, with a similar genetic background, an increased level of a homogentisate prenyl transferase protein or fragment thereof.

In some embodiments, the levels of one or more products of the tocopherol biosynthesis pathway, including any one or more of tocopherols, α-tocopherols, γ-tocopherols, δ-tocopherols, β-tocopherols, tocotrienols, α-tocotrienols, γ-tocotrienols, δ-tocotrienols, β-tocotrienols are increased by greater than about 10%, or more preferably greater than about 25%, 35%, 50%, 75%, 80%, 90%, 100%, 150%, 200%, 1,000%, 2,000%, or 2,500%. The levels of products may be increased throughout an organism such as a plant or localized in one or more specific organs or tissues of the organism. For example, the levels of products may be increased in one or more of the tissues and organs of a plant including without limitation: roots, tubers, stems, leaves, stalks, fruit, berries, nuts, bark, pods, seeds and flowers. A preferred organ is a seed.

In some embodiments, the levels of one or more products of the tocopherol biosynthesis pathway, including any one or more of tocopherols, α-tocopherols, γ-tocopherols, δ-tocopherols, β-tocopherols, tocotrienols, α-tocotrienols, γ-tocotrienols, δ-tocotrienols, β-tocotrienols are increased so that they constitute greater than about 10%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the total tocopherol content of the organism or tissue. The levels of products may be increased throughout an organism such as a plant or localized in one or more specific organs or tissues of the organism. For example, the levels of products may be increased in one or more of the tissues and organs of a plant including without limitation: roots, tubers, stems, leaves, stalks, fruit, berries, nuts, bark, pods, seeds and flowers. A preferred organ is a seed.

In a preferred embodiment, expression of enzymes involved in tocopherol, tocotrienol or plastoquinol synthesis in the seed will result in an increase in γ-tocopherol levels due to the absence of significant levels of GMT activity in those tissues. In another preferred embodiment, expression of enzymes involved in tocopherol, tocotrienol, or plastoquinol synthesis in photosyhthetic tissues will result in an increase in α-tocopherol due to the higher levels of GMT activity in those tissues relative to the same activity in seed tissue.

In another preferred embodiment, the expression of enzymes involved in tocopherol, tocotrienol, or plastoquinol synthesis in the seed will result in an increase in the total tocopherol, tocotrienol, or plastoquinol level in the plant.

In some embodiments, the levels of tocopherols or a species such as α-tocopherol may be altered. In some embodiments, the levels of tocotrienols may be altered. Such alteration can be compared to a plant with a similar background.

In another embodiment, either the α-tocopherol level, α-tocotrienol level, or both of plants that natively produce high levels of either α-tocopherol, α-tocotrienol or both (*e.g.*, sunflowers), can be increased by the introduction of a gene coding for a homogentisate prenyl transferase enzyme.

In a preferred aspect, a similar genetic background is a background where the organisms being compared share about 50% or greater of their nuclear genetic material. In a more preferred aspect a similar genetic background is a background where the organisms being compared share about 75% or greater, even more preferably about 90% or greater of their nuclear genetic material. In another even more preferable aspect, a similar genetic background is a background where the organisms being compared are plants, and the plants are isogenic except for any genetic material originally introduced using plant transformation techniques.

In another preferred embodiment, expression or overexpression of a polypeptide of the present invention in a transformed plant may provide tolerance to a variety of stress, *e.g.* oxidative stress tolerance such as to oxygen or ozone, UV tolerance, cold tolerance, or fungal/microbial pathogen tolerance.

As used herein in a preferred aspect, a tolerance or resistance to stress is determined by the ability of a plant, when challenged by a stress such as cold to produce a plant having a higher yield than one without such tolerance or resistance to stress. In a particularly preferred aspect of the present invention, the tolerance or resistance to stress is measured relative to a plant with a similar genetic background to the tolerant or resistance plant except that the plant reduces the expression, expresses, or over expresses a protein or fragment thereof of the present invention.

Exogenous genetic material may be transferred into a host cell by the use of a DNA vector or construct designed for such a purpose. Design of such a vector is generally within

the skill of the art (*see*, *Plant Molecular Biology: A Laboratory Manual*, Clark (ed.), Springer, NY (1997)).

A construct or vector may include a plant promoter to express the polypeptide of choice. In a preferred embodiment, any nucleic acid molecules described herein can be operably linked to a promoter region which functions in a plant cell to cause the production of an mRNA molecule. For example, any promoter that functions in a plant cell to cause the production of an mRNA molecule, such as those promoters described herein, without limitation, can be used. In a preferred embodiment, the promoter is a plant promoter.

A number of promoters that are active in plant cells have been described in the literature. These include the nopaline synthase (NOS) promoter (Ebert *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 84:5745-5749 (1987)), the octopine synthase (OCS) promoter (which is carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*), the caulimovirus promoters such as the cauliflower mosaic virus (CaMV) 19S promoter (Lawton *et al.*, *Plant Mol. Biol.*, 9:315-324 (1987)) and the CaMV 35S promoter (Odell *et al.*, *Nature*, 313:810-812 (1985)), the figwort mosaic virus 35S-promoter, the light-inducible promoter from the small subunit of ribulose-1,5-bis-phosphate carboxylase (ssRUBISCO), the Adh promoter (Walker *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 84:6624-6628 (1987)), the sucrose synthase promoter (Yang *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 87:4144-4148 (1990)), the R gene complex promoter (Chandler *et al.*, *The Plant Cell*, 1:1175-1183 (1989)) and the chlorophyll a/b binding protein gene promoter, *etc.* These promoters have been used to create DNA constructs that have been expressed in plants; *see*, *e.g.*, WO 84/02913. The CaMV 35S promoters are preferred for use in plants. Promoters known or found to cause transcription of DNA in plant cells can be used in the present invention.

For the purpose of expression in source tissues of the plant, such as the leaf, seed, root or stem, it is preferred that the promoters utilized have relatively high expression in these specific tissues. Tissue-specific expression of a protein of the present invention is a particularly preferred embodiment. For this purpose, one may choose from a number of promoters for genes with tissue- or cell-specific or enhanced expression. Examples of such promoters reported in the literature include the chloroplast glutamine synthetase GS2 promoter from pea (Edwards *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 87:3459-3463 (1990)), the chloroplast fructose-1,6-biphosphatase (FBPase) promoter from wheat (Lloyd *et al.*, *Mol. Gen. Genet.*, 225:209-216 (1991)), the nuclear photosynthetic ST-LS1 promoter from potato

(Stockhaus *et al.*, *EMBO J.*, 8:2445-2451 (1989)), the serine/threonine kinase (PAL) promoter and the glucoamylase (CHS) promoter from *Arabidopsis thaliana*. Also, reported to be active in photosynthetically active tissues are the ribulose-1,5-bisphosphate carboxylase (RbcS) promoter from eastern larch (*Larix laricina*), the promoter for the *cab* gene, cab6, from pine (Yamamoto *et al.*, *Plant Cell Physiol.*, 35:773-778 (1994)), the promoter for the Cab-1 gene from wheat (Fejes *et al.*, *Plant Mol. Biol.*, 15:921-932 (1990)), the promoter for the CAB-1 gene from spinach (Lubberstedt *et al.*, *Plant Physiol.*, 104:997-1006 (1994)), the promoter for the cab1R gene from rice (Luan *et al.*, *Plant Cell.*, 4:971-981 (1992)), the pyruvate, orthophosphate dikinase (PPDK) promoter from corn (Matsuoka *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 90:9586-9590 (1993)), the promoter for the tobacco Lhcb1*2 gene (Cerdan *et al.*, *Plant Mol. Biol.*, 33:245-255 (1997)), the *Arabidopsis thaliana* SUC2 sucrose-H+ symporter promoter (Truernit *et al.*, *Planta.*, 196:564-570 (1995)) and the promoter for the thylakoid membrane proteins from spinach (*psa*D, *psa*F, *psa*E, PC, FNR, *atp*C, *atp*D, *cab*, *rbc*S). Other promoters for the chlorophyll a/b-binding proteins may also be utilized in the present invention, such as the promoters for *Lhc*B gene and *Psb*P gene from white mustard (*Sinapis alba*; Kretsch *et al.*, *Plant Mol. Biol.*, 28:219-229 (1995)).

For the purpose of expression in sink tissues of the plant, such as the tuber of the potato plant, the fruit of tomato, or the seed of corn, wheat, rice and barley, it is preferred that the promoters utilized in the present invention have relatively high expression in these specific tissues. A number of promoters for genes with tuber-specific or tuber-enhanced expression are known, including the class I patatin promoter (Bevan *et al.*, *EMBO J.*, 8:1899-1906 (1986); Jefferson *et al.*, *Plant Mol. Biol.*, 14:995-1006 (1990)), the promoter for the potato tuber ADPGPP genes, both the large and small subunits, the sucrose synthase promoter (Salanoubat and Belliard, *Gene*, 60:47-56 (1987), Salanoubat and Belliard, *Gene*, 84:181-185 (1989)), the promoter for the major tuber proteins including the 22 kd protein complexes and protease inhibitors (Hannapel, *Plant Physiol.*, 101:703-704 (1993)), the promoter for the granule-bound starch synthase gene (GBSS) (Visser *et al.*, *Plant Mol. Biol.*, 17:691-699 (1991)) and other class I and II patatins promoters (Koster-Topfer *et al.*, *Mol. Gen. Genet.*, 219:390-396 (1989); Mignery *et al.*, *Gene.*, 62:27-44 (1988)).

Other promoters can also be used to express a polypeptide in specific tissues, such as seeds or fruits. Indeed, in a preferred embodiment, the promoter used is a seed specific promoter. Examples of such promoters include the 5' regulatory regions from such genes as

napin (Kridl *et al., Seed Sci. Res.*, 1:209:219 (1991)), phaseolin (Bustos *et al., Plant Cell*, 1(9):839-853 (1989)), soybean trypsin inhibitor (Riggs *et al., Plant Cell*, 1(6):609-621 (1989)), ACP (Baerson *et al., Plant Mol. Biol.*, 22(2):255-267 (1993)), stearoyl-ACP desaturase (Slocombe *et al., Plant Physiol.*, 104(4):167-176 (1994)), soybean α' subunit of β-conglycinin (soy 7s, (Chen *et al., Proc. Natl. Acad. Sci.*, 83:8560-8564 (1986))), and oleosin (*see*, for example, Hong *et al., Plant Mol. Biol.*, 34(3):549-555 (1997)). Further examples include the promoter for β-conglycinin (Chen *et al., Dev. Genet.*, 10:112-122 (1989)). Also included are the zeins, which are a group of storage proteins found in corn endosperm. Genomic clones for zein genes have been isolated (Pedersen *et al., Cell*, 29:1015-1026 (1982), and Russell *et al., Transgenic Res.*, 6(2):157-168) and the promoters from these clones, including the 15 kD, 16 kD, 19 kD, 22 kD, 27 kD and genes, could also be used. Other promoters known to function, for example, in corn include the promoters for the following genes: *waxy, Brittle, Shrunken 2*, Branching enzymes I and II, starch synthases, debranching enzymes, oleosins, glutelins and sucrose synthases. A particularly preferred promoter for corn endosperm expression is the promoter for the glutelin gene from rice, more particularly the Osgt-1 promoter (Zheng *et al., Mol. Cell Biol.*, 13:5829-5842 (1993)). Examples of promoters suitable for expression in wheat include those promoters for the ADPglucose pyrosynthase (ADPGPP) subunits, the granule bound and other starch synthase, the branching and debranching enzymes, the embryogenesis-abundant proteins, the gliadins and the glutenins. Examples of such promoters in rice include those promoters for the ADPGPP subunits, the granule bound and other starch synthase, the branching enzymes, the debranching enzymes, sucrose synthases and the glutelins. A particularly preferred promoter is the promoter for rice glutelin, Osgt-1. Examples of such promoters for barley include those for the ADPGPP subunits, the granule bound and other starch synthase, the branching enzymes, the debranching enzymes, sucrose synthases, the hordeins, the embryo globulins and the aleurone specific proteins. A preferred promoter for expression in the seed is a napin promoter. Another preferred promoter for expression is an Arcelin 5 promoter.

Root specific promoters may also be used. An example of such a promoter is the promoter for the acid chitinase gene (Samac *et al., Plant Mol. Biol.*, 25:587-596 (1994)). Expression in root tissue could also be accomplished by utilizing the root specific subdomains of the CaMV35S promoter that have been identified (Lam *et al., Proc. Natl. Acad. Sci.*

(U.S.A.), 86:7890-7894 (1989)). Other root cell specific promoters include those reported by Conkling *et al., Plant Physiol.*, 93:1203-1211 (1990).

Other preferred promoters include 7Sα' (Beachy *et al.*, *EMBO J.*, 4:3047 (1985); Schuler *et al.*, *Nucleic Acid Res.*, 10(24):8225-8244 (1982)); USP 88 and enhanced USP 88, and 7Sα, (U.S. Patent No. 6,825,398).

Additional promoters that may be utilized are described, for example, in U.S. Patents 5,378,619; 5,391,725; 5,428,147; 5,447,858; 5,608,144; 5,608,144; 5,614,399; 5,633,441; 5,633,435; and 4,633,436. In addition, a tissue specific enhancer may be used (Fromm *et al.*, *The Plant Cell*, 1:977-984 (1989)).

Constructs or vectors may also include, with the coding region of interest, a nucleic acid sequence that acts, in whole or in part, to terminate transcription of that region. A number of such sequences have been isolated, including the Tr7 3' sequence and the NOS 3' sequence (Ingelbrecht *et al.*, *The Plant Cell*, 1:671-680 (1989); Bevan *et al.*, *Nucleic Acids Res.*, 11:369-385 (1983)). Regulatory transcript termination regions can be provided in plant expression constructs of this present invention as well. Transcript termination regions can be provided by the DNA sequence encoding the gene of interest or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region that is naturally associated with the transcript initiation region. The skilled artisan will recognize that any convenient transcript termination region that is capable of terminating transcription in a plant cell can be employed in the constructs of the present invention.

A vector or construct may also include regulatory elements. Examples of such include the Adh intron 1 (Callis *et al.*, *Genes and Develop.*, 1:1183-1200 (1987)), the sucrose synthase intron (Vasil *et al.*, *Plant Physiol.*, 91:1575-1579 (1989)) and the TMV omega element (Gallie *et al.*, *The Plant Cell*, 1:301-311 (1989)). These and other regulatory elements may be included when appropriate.

A vector or construct may also include a selectable marker. Selectable markers may also be used to select for plants or plant cells that contain the exogenous genetic material. Examples of such include, but are not limited to: a *neo* gene (Potrykus *et al.*, *Mol. Gen. Genet.*, 199:183-188 (1985)), which codes for kanamycin resistance and can be selected for using kanamycin, RptII, G418, hpt *etc.*; a bar gene which codes for bialaphos resistance; a

mutant EPSP synthase gene (Hinchee *et al.*, *Bio/Technology*, 6:915-922 (1988); Reynaerts *et al.*, Selectable and Screenable Markers. In: Gelvin and Schilperoort, Plant Molecular Biology Manual, Kluwer, Dordrecht (1988); Reynaerts *et al.*, Selectable and Screenable Markers. In: Gelvin and Schilperoort, Plant Molecular Biology Manual, Kluwer, Dordrecht (1988)), *aad*A (Jones *et al., Mol. Gen. Genet.* (1987)), which encodes glyphosate resistance; a nitrilase gene which confers resistance to bromoxynil (Stalker *et al., J. Biol. Chem.*, 263:6310-6314 (1988)); a mutant acetolactate synthase gene (ALS) which confers imidazolinone or sulphonylurea resistance (EP 0 154 204 (Sept. 11, 1985)), ALS (D'Halluin *et al., Bio/Technology*, 10:309-314 (1992)), and a methotrexate resistant DHFR gene (Thillet *et al., J. Biol. Chem.*, 263:12500-12508 (1988)).

A vector or construct may also include a transit peptide. Incorporation of a suitable chloroplast transit peptide may also be employed (EP 0 218 571). Translational enhancers may also be incorporated as part of the vector DNA. DNA constructs could contain one or more 5' non-translated leader sequences, which may serve to enhance expression of the gene products from the resulting mRNA transcripts. Such sequences may be derived from the promoter selected to express the gene or can be specifically modified to increase translation of the mRNA. Such regions may also be obtained from viral RNAs, from suitable eukaryotic genes, or from a synthetic gene sequence. For a review of optimizing expression of transgenes, see Koziel *et al., Plant Mol. Biol.*, 32:393-405 (1996). A preferred transit peptide is CTP1.

A vector or construct may also include a screenable marker. Screenable markers may be used to monitor expression. Exemplary screenable markers include: a β-glucuronidase or *uid*A gene (GUS) which encodes an enzyme for which various chromogenic substrates are known (Jefferson, *Plant Mol. Biol, Rep.*, 5:387-405 (1987); Jefferson *et al., EMBO J.*, 6:3901-3907 (1987)); an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta *et al., Stadler Symposium*, 11:263-282 (1988)); a β-lactamase gene (Sutcliffe *et al., Proc. Natl. Acad. Sci. (U.S.A.)*, 75:3737-3741 (1978)), a gene which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); a luciferase gene (Ow *et al., Science*, 234:856-859 (1986)); a *xyl*E gene (Zukowsky *et al., Proc. Natl. Acad. Sci. (U.S.A.)*, 80:1101-1105 (1983)) which encodes a catechol dioxygenase that can convert chromogenic catechols; an α-amylase gene (Ikatu *et al., Bio/Technol.*, 8:241-242 (1990)); a

tyrosinase gene (Katz *et al.*, *J. Gen. Microbiol.*, 129:2703-2714 (1983)) which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to melanin; an α-galactosidase, which will turn a chromogenic α-galactose substrate.

Included within the terms "selectable or screenable marker genes" are also genes that encode a secretable marker whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers that encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes that can be detected catalytically. Secretable proteins fall into a number of classes, including small, diffusible proteins that are detectable, (*e.g.*, by ELISA), small active enzymes that are detectable in extracellular solution (*e.g.*, α-amylase, β-lactamase, phosphinothricin transferase), or proteins that are inserted or trapped in the cell wall (such as proteins that include a leader sequence such as that found in the expression unit of extension or tobacco PR-S). Other possible selectable and/or screenable marker genes will be apparent to those of skill in the art.

There are many methods for introducing transforming nucleic acid molecules into plant cells. Suitable methods are believed to include virtually any method by which nucleic acid molecules may be introduced into a cell, such as by *Agrobacterium* infection or direct delivery of nucleic acid molecules such as, for example, by PEG-mediated transformation, by electroporation or by acceleration of DNA coated particles, and the like. (Potrykus, *Ann. Rev. Plant Physiol. Plant Mol. Biol.*, 42:205-225 (1991); Vasil, *Plant Mol. Biol.*, 25:925-937 (1994)). For example, electroporation has been used to transform corn protoplasts (Fromm *et al.*, *Nature*, 312:791-793 (1986)).

Other vector systems suitable for introducing transforming DNA into a host plant cell include but are not limited to binary artificial chromosome (BIBAC) vectors (Hamilton *et al.*, *Gene*, 200:107-116 (1997)); and transfection with RNA viral vectors (Della-Cioppa *et al.*, *Ann. N.Y. Acad. Sci.* (1996), 792 (Engineering Plants for Commercial Products and Applications, 57-61). Additional vector systems also include plant selectable YAC vectors such as those described in Mullen *et al.*, *Molecular Breeding*, 4:449-457 (1988).

Technology for introduction of DNA into cells is well known to those of skill in the art. Four general methods for delivering a gene into cells have been described: (1) chemical methods (Graham and van der Eb, *Virology*, 54:536-539 (1973)); (2) physical methods such as microinjection (Capecchi, *Cell*, 22:479-488 (1980)), electroporation (Wong and Neumann, *Biochem. Biophys. Res. Commun.*, 107:584-587 (1982); Fromm *et al.*, *Proc. Natl. Acad. Sci.*

(U.S.A.), 82:5824-5828 (1985); U.S. Patent 5,384,253); the gene gun (Johnston and Tang, *Methods Cell Biol.*, 43:353-365 (1994)); and vacuum infiltration (Bechtold *et al.*, *C.R. Acad. Sci. Paris, Life Sci.*, 316:1194-1199 (1993)); (3) viral vectors (Clapp, *Clin. Perinatol.*, 20:155-168 (1993); Lu *et al.*, *J. Exp. Med.*, 178:2089-2096 (1993); Eglitis and Anderson, *Biotechniques*, 6:608-614 (1988)); and (4) receptor-mediated mechanisms (Curiel *et al.*, *Hum. Gen. Ther.*, 3:147-154 (1992), Wagner *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 89:6099-6103 (1992)).

Acceleration methods that may be used include, for example, microprojectile bombardment and the like. One example of a method for delivering transforming nucleic acid molecules into plant cells is microprojectile bombardment. This method has been reviewed by Yang and Christou (eds.), *Particle Bombardment Technology for Gene Transfer*, Oxford Press, Oxford, England (1994). Non-biological particles (microprojectiles) may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum and the like.

A particular advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly transforming monocots, is that neither the isolation of protoplasts (Cristou *et al.*, *Plant Physiol.*, 87:671-674 (1988)) nor the susceptibility to *Agrobacterium* infection is required. An illustrative embodiment of a method for delivering DNA into corn cells by acceleration is a biolistics α-particle delivery system, which can be used to propel particles coated with DNA through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with corn cells cultured in suspension. Gordon-Kamm *et al.*, describes the basic procedure for coating tungsten particles with DNA (Gordon-Kamm *et al.*, *Plant Cell*, 2:603-618 (1990)). The screen disperses the tungsten nucleic acid particles so that they are not delivered to the recipient cells in large aggregates. A particle delivery system suitable for use with the present invention is the helium acceleration PDS-1000/He gun, which is available from Bio-Rad Laboratories (Bio-Rad, Hercules, CA) (Sanford *et al.*, *Technique*, 3:3-16 (1991)).

For the bombardment, cells in suspension may be concentrated on filters. Filters containing the cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate. If desired, one or more screens are also positioned between the gun and the cells to be bombarded.

Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate. If desired, one or more screens are also positioned between the acceleration device and the cells to be bombarded. Through the use of techniques set forth herein one may obtain 1000 or more loci of cells transiently expressing a marker gene. The number of cells in a focus that express the exogenous gene product 48 hours post-bombardment often ranges from one to ten, and average one to three.

In bombardment transformation, one may optimize the pre-bombardment culturing conditions and the bombardment parameters to yield the maximum numbers of stable transformants. Both the physical and biological parameters for bombardment are important in this technology. Physical factors are those that involve manipulating the DNA/microprojectile precipitate or those that affect the flight and velocity of either the macro- or microprojectiles. Biological factors include all steps involved in manipulation of cells before and immediately after bombardment, the osmotic adjustment of target cells to help alleviate the trauma associated with bombardment and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. It is believed that pre-bombardment manipulations are especially important for successful transformation of immature embryos.

In another alternative embodiment, plastids can be stably transformed. Methods disclosed for plastid transformation in higher plants include the particle gun delivery of DNA containing a selectable marker and targeting of the DNA to the plastid genome through homologous recombination (Svab *et al., Proc. Natl. Acad. Sci. (U.S.A.)*, 87:8526-8530 (1990); Svab and Maliga, *Proc. Natl. Acad. Sci. (U.S.A.)*, 90:913-917 (1993); Staub and Maliga, *EMBO J.*, 12:601-606 (1993); U.S. Patents 5,451,513 and 5,545,818).

Accordingly, it is contemplated that one may wish to adjust various aspects of the bombardment parameters in small scale studies to fully optimize the conditions. One may particularly wish to adjust physical parameters such as gap distance, flight distance, tissue distance and helium pressure. One may also minimize the trauma reduction factors by modifying conditions that influence the physiological state of the recipient cells and which may therefore influence transformation and integration efficiencies. For example, the osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells may be

adjusted for optimum transformation. The execution of other routine adjustments will be known to those of skill in the art in light of the present disclosure.

Agrobacterium-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. *See*, for example, the methods described by Fraley *et al., Bio/Technology*, 3:629-635 (1985) and Rogers *et al., Methods Enzymol.*, 153:253-277 (1987). Further, the integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences and intervening DNA is usually inserted into the plant genome as described (Spielmann *et al., Mol. Gen. Genet.*, 205:34 (1986)).

Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as described (Klee *et al.*, In: *Plant DNA Infectious Agents*, Hohn and Schell (eds.), Springer-Verlag, NY, pp. 179-203 (1985)). Moreover, technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes (Rogers *et al., Methods Enzymol.*, 153:253-277 (1987)). In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

A transgenic plant formed using *Agrobacterium* transformation methods typically contains a single gene on one chromosome. Such transgenic plants can be referred to as being heterozygous for the added gene. More preferred is a transgenic plant that is homozygous for the added structural gene; *i.e.*, a transgenic plant that contains two added genes, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) an independent segregant, transgenic plant that

contains a single added gene, germinating some of the seed produced and analyzing the resulting plants produced for the gene of interest.

It is also to be understood that two different transgenic plants can also be mated to produce offspring that contain two independently segregating, exogenous genes. Selfing of appropriate progeny can produce plants that are homozygous for both added, exogenous genes that encode a polypeptide of interest. Back-crossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated, as is vegetative propagation.

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation and combinations of these treatments (see, *for example*, Potrykus *et al., Mol. Gen. Genet.*, 205:193-200 (1986); Lorz *et al., Mol. Gen. Genet.*, 199:178 (1985); Fromm *et al., Nature*, 319:791 (1986); Uchimiya *et al., Mol. Gen. Genet.*, 204:204 (1986); Marcotte *et al., Nature*, 335:454-457 (1988)).

Application of these systems to different plant strains depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts are described (Fujimura *et al., Plant Tissue Culture Letters*, 2:74 (1985); Toriyama *et al., Theor. Appl. Genet.*, 205:34 (1986); Yamada *et al., Plant Cell Rep.*, 4:85 (1986); Abdullah *et al., Biotechnology*, 4:1087 (1986)).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil, *Biotechnology*, 6:397 (1988)). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil *et al., Bio/Technology*, 10:667 (1992)).

Using the latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al., Nature*, 328:70 (1987); Klein *et al., Proc. Natl. Acad. Sci. (U.S.A.)*, 85:8502-8505 (1988); McCabe *et al., Bio/Technology*, 6:923 (1988)). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants.

Other methods of cell transformation can also be used and include but are not limited to introduction of DNA into plants by direct DNA transfer into pollen (Hess *et al., Intern Rev. Cytol.*, 107:367 (1987); Luo *et al., Plant Mol Biol. Reporter*, 6:165 (1988)), by direct injection of DNA into reproductive organs of a plant (Pena *et al., Nature*, 325:274 (1987)), or by direct

injection of DNA into the cells of immature embryos followed by the rehydration of desiccated embryos (Neuhaus *et al., Theor. Appl. Genet.*, 75:30 (1987)).

The regeneration, development and cultivation of plants from single plant protoplast transformants or from various transformed explants is well known in the art (Weissbach and Weissbach, In: *Methods for Plant Molecular Biology*, Academic Press, San Diego, CA, (1988)). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil.

The development or regeneration of plants containing the foreign, exogenous gene that encodes a protein of interest is well known in the art. Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important lines. Conversely, pollen from plants of these important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

There are a variety of methods for the regeneration of plants from plant tissue. The particular method of regeneration will depend on the starting plant tissue and the particular plant species to be regenerated.

Methods for transforming dicots, primarily by use of *Agrobacterium tumefaciens* and obtaining transgenic plants have been published for cotton (U.S. Patents 5,004,863; 5,159,135; and 5,518,908); soybean (U.S. Patents 5,569,834 and 5,416,011; McCabe *et al., Biotechnology*, 6:923 (1988); Christou *et al., Plant Physiol.*, 87:671-674 (1988)); *Brassica* (U.S. Patent 5,463,174); peanut (Cheng *et al., Plant Cell Rep.*, 15:653-657 (1996), McKently *et al., Plant Cell Rep.*, 14:699-703 (1995)); papaya; pea (Grant *et al., Plant Cell Rep.*, 15:254-258 (1995)); and *Arabidopsis thaliana* (Bechtold *et al., C.R. Acad. Sci. Paris, Life Sci.*, 316:1194-1199 (1993)). The latter method for transforming *Arabidopsis thaliana* is commonly called "dipping" or vacuum infiltration or germplasm transformation.

Transformation of monocotyledons using electroporation, particle bombardment and *Agrobacterium* have also been reported. Transformation and plant regeneration have been achieved in asparagus (Bytebier *et al., Proc. Natl. Acad. Sci. (U.S.A.)*, 84:5354 (1987)); barley

(Wan and Lemaux, *Plant Physiol*, 104:37 (1994)); corn (Rhodes *et al., Science*, 240:204 (1988); Gordon-Kamm *et al., Plant Cell*, 2:603-618 (1990); Fromm *et al., Bio/Technology*, 8:833 (1990); Koziel *et al., Bio/Technology*, 11:194 (1993); Armstrong *et al., Crop Science*, 35:550-557 (1995)); oat (Somers *et al., Bio/Technology*, 10:1589 (1992)); orchard grass (Horn *et al., Plant Cell Rep.*, 7:469 (1988)); rice (Toriyama *et al., Theor Appl. Genet.*, 205:34 (1986); Part *et al., Plant Mol. Biol.*, 32:1135-1148 (1996); Abedinia *et al., Aust. J. Plant Physiol.*, 24:133-141 (1997); Zhang and Wu, *Theor. Appl. Genet.*, 76:835 (1988); Zhang *et al., Plant Cell Rep.*, 7:379 (1988); Battraw and Hall, *Plant Sci.*, 86:191-202 (1992); Christou *et al., Bio/Technology*, 9:957 (1991)); rye (De la Pena *et al., Nature*, 325:274 (1987)); sugarcane (Bower and Birch, *Plant J.*, 2:409 (1992)); tall fescue (Wang *et al., Bio/Technology*, 10:691 (1992)); and wheat (Vasil *et al., Bio/Technology*, 10:667 (1992); U.S. Patent 5,631,152).

Assays for gene expression based on the transient expression of cloned nucleic acid constructs have been developed by introducing the nucleic acid molecules into plant cells by polyethylene glycol treatment, electroporation, or particle bombardment (Marcotte *et al., Nature*, 335:454-457 (1988); Marcotte *et al., Plant Cell*, 1:523-532 (1989); McCarty *et al., Cell*, 66:895-905 (1991); Hattori *et al., Genes Dev.*, 6:609-618 (1992); Goff *et al., EMBO J.*, 9:2517-2522 (1990)). Transient expression systems may be used to functionally dissect gene constructs (*see generally*, Mailga *et al., Methods in Plant Molecular Biology,* Cold Spring Harbor Press, NY (1995)).

Any of the nucleic acid molecules of the present invention may be introduced into a plant cell in a permanent or transient manner in combination with other genetic elements such as vectors, promoters, enhancers, *etc.* Further, any of the nucleic acid molecules of the present invention may be introduced into a plant cell in a manner that allows for expression or overexpression of the protein or fragment thereof encoded by the nucleic acid molecule.

Cosuppression is the reduction in expression levels, usually at the level of RNA, of a particular endogenous gene or gene family by the expression of a homologous sense construct that is capable of transcribing mRNA of the same strandedness as the transcript of the endogenous gene (Napoli *et al., Plant Cell*, 2:279-289 (1990); van der Krol *et al., Plant Cell*, 2:291-299 (1990)). Cosuppression may result from stable transformation with a single copy nucleic acid molecule that is homologous to a nucleic acid sequence found with the cell (Prolls and Meyer, *Plant J.*, 2:465-475 (1992)) or with multiple copies of a nucleic acid

molecule that is homologous to a nucleic acid sequence found with the cell (Mittlesten *et al., Mol. Gen. Genet.*, 244:325-330 (1994)). Genes, even though different, linked to homologous promoters may result in the cosuppression of the linked genes (Vaucheret, *C.R. Acad. Sci. III*, 316:1471-1483 (1993); Flavell, *Proc. Natl. Acad. Sci. (U.S.A.)*, 91:3490-3496 (1994)); van Blokland *et al., Plant J.*, 6:861-877 (1994); Jorgensen, *Trends Biotechnol.*, 8:340-344 (1990); Meins and Kunz, In: *Gene Inactivation and Homologous Recombination in Plants*, Paszkowski (ed.), pp. 335-348, Kluwer Academic, Netherlands (1994)).

It is understood that one or more of the nucleic acids of the present invention may be introduced into a plant cell and transcribed using an appropriate promoter with such transcription resulting in the cosuppression of an endogenous protein.

Antisense approaches are a way of preventing or reducing gene function by targeting the genetic material (Mol *et al., FEBS Lett.*, 268:427-430 (1990)). The objective of the antisense approach is to use a sequence complementary to the target gene to block its expression and create a mutant cell line or organism in which the level of a single chosen protein is selectively reduced or abolished. Antisense techniques have several advantages over other "reverse genetic" approaches. The site of inactivation and its developmental effect can be manipulated by the choice of promoter for antisense genes or by the timing of external application or microinjection. Antisense can manipulate its specificity by selecting either unique regions of the target gene or regions where it shares homology to other related genes (Hiatt *et al.,* In: *Genetic Engineering,* Setlow (ed.), Vol. 11, New York: Plenum 49-63 (1989)).

Antisense RNA techniques involve introduction of RNA that is complementary to the target mRNA into cells, which results in specific RNA:RNA duplexes being formed by base pairing between the antisense substrate and the target mRNA (Green *et al., Annu. Rev. Biochem.*, 55:569-597 (1986)). Under one embodiment, the process involves the introduction and expression of an antisense gene sequence. Such a sequence is one in which part or all of the normal gene sequences are placed under a promoter in inverted orientation so that the "wrong" or complementary strand is transcribed into a noncoding antisense RNA that hybridizes with the target mRNA and interferes with its expression (Takayama and Inouye, *Crit. Rev. Biochem. Mol. Biol.*, 25:155-184 (1990)). An antisense vector is constructed by standard procedures and introduced into cells by transformation, transfection, electroporation, microinjection, infection, *etc.* The type of transformation and choice of vector will determine

whether expression is transient or stable. The promoter used for the antisense gene may influence the level, timing, tissue, specificity, or inducibility of the antisense inhibition.

It is understood that the activity of a protein in a plant cell may be reduced or depressed by growing a transformed plant cell containing a nucleic acid molecule whose non-transcribed strand encodes a protein or fragment thereof. A preferred protein whose activity can be reduced or depressed, by any method, is a homogentisate prenyl transferase.

Posttranscriptional gene silencing (PTGS) can result in virus immunity or gene silencing in plants. PTGS is induced by dsRNA and is mediated by an RNA-dependent RNA polymerase, present in the cytoplasm, which requires a dsRNA template. The dsRNA is formed by hybridization of complementary transgene mRNAs or complementary regions of the same transcript. Duplex formation can be accomplished by using transcripts from one sense gene and one antisense gene colocated in the plant genome, a single transcript that has self-complementarity, or sense and antisense transcripts from genes brought together by crossing. The dsRNA-dependent RNA polymerase makes a complementary strand from the transgene mRNA and RNAse molecules attach to this complementary strand (cRNA). These cRNA-RNase molecules hybridize to the endogene mRNA and cleave the single-stranded RNA adjacent to the hybrid. The cleaved single-stranded RNAs are further degraded by other host RNases because one will lack a capped 5' end and the other will lack a poly (A) tail (Waterhouse *et al., PNAS*, 95:13959-13964 (1998)).

It is understood that one or more of the nucleic acids of the present invention may be introduced into a plant cell and transcribed using an appropriate promoter with such transcription resulting in the posttranscriptional gene silencing of an endogenous transcript.

Antibodies have been expressed in plants (Hiatt *et al., Nature*, 342:76-78 (1989); Conrad and Fielder, *Plant Mol. Biol.*, 26:1023-1030 (1994)). Cytoplasmic expression of a scFv (single-chain Fv antibody) has been reported to delay infection by artichoke mottled crinkle virus. Transgenic plants that express antibodies directed against endogenous proteins may exhibit a physiological effect (Philips *et al., EMBO J.*, 16:4489-4496 (1997); Marion-Poll, *Trends in Plant Science*, 2:447-448 (1997)). For example, expressed anti-abscisic antibodies have been reported to result in a general perturbation of seed development (Philips *et al., EMBO J.*, 16:4489-4496 (1997)).

Antibodies that are catalytic may also be expressed in plants (abzymes). The principle behind abzymes is that since antibodies may be raised against many molecules, this

recognition ability can be directed toward generating antibodies that bind transition states to force a chemical reaction forward (Persidas, *Nature Biotechnology*, 15:1313-1315 (1997); Baca *et al., Ann. Rev. Biophys. Biomol. Struct.*, 26:461-493 (1997)). The catalytic abilities of abzymes may be enhanced by site directed mutagenesis. Examples of abzymes are, for example, set forth in U.S. Patents: 5,658,753; 5,632,990; 5,631,137; 5,602,015; 5,559,538; 5,576,174; 5,500,358; 5,318,897; 5,298,409; 5,258,289; and 5,194,585.

It is understood that any of the antibodies of the present invention may be expressed in plants and that such expression can result in a physiological effect. It is also understood that any of the expressed antibodies may be catalytic.

The present invention also provides for parts of the plants, particularly reproductive or storage parts, of the present invention. Plant parts, without limitation, include seed, endosperm, ovule and pollen. In a particularly preferred embodiment of the present invention, the plant part is a seed. In one embodiment the seed is a constituent of animal feed.

In another embodiment, the plant part is a fruit, more preferably a fruit with enhanced shelf life. In another preferred embodiment, the fruit has increased levels of a tocopherol. In another preferred embodiment, the fruit has increased levels of a tocotrienol.

The present invention also provides a container of over about 10,000, more preferably about 20,000, and even more preferably about 40,000 seeds where over about 10%, more preferably 25%, more preferably 50%, and even more preferably 75% or 90% of the seeds are seeds derived from a plant of the present invention.

The present invention also provides a container of over about 10 kg, more preferably 25 kg, and even more preferably 50 kg seeds where over about 10%, more preferably 25%, more preferably 50%, and even more preferably 75% or 90% of the seeds are seeds derived from a plant of the present invention.

Any of the plants or parts thereof of the present invention may be processed to produce a feed, meal, protein, or oil preparation, including oil preparations high in total tocopherol content and oil preparations high in any one or more of each tocopherol component listed herein. A particularly preferred plant part for this purpose is a seed. In a preferred embodiment the feed, meal, protein or oil preparation is designed for livestock animals or humans, or both. Methods to produce feed, meal, protein and oil preparations are known in the art. *See*, for example, U.S. Patents 4,957,748; 5,100,679; 5,219,596; 5,936,069; 6,005,076; 6,146,669; and 6,156,227. In a preferred embodiment, the protein preparation is a

high protein preparation. Such a high protein preparation preferably has a protein content of greater than about 5% w/v, more preferably 10% w/v, and even more preferably 15% w/v. In a preferred oil preparation, the oil preparation is a high oil preparation with an oil content derived from a plant or part thereof of the present invention of greater than about 5% w/v, more preferably 10% w/v, and even more preferably 15% w/v. In a preferred embodiment the oil preparation is a liquid and of a volume greater than about 1, 5, 10, or 50 liters. The present invention provides for oil produced from plants of the present invention or generated by a method of the present invention. Such an oil may exhibit enhanced oxidative stability. Also, such oil may be a minor or major component of any resultant product. Moreover, such oil may be blended with other oils. In a preferred embodiment, the oil produced from plants of the present invention or generated by a method of the present invention constitutes greater than about 0.5%, 1%, 5%, 10%, 25%, 50%, 75%, or 90% by volume or weight of the oil component of any product. In another embodiment, the oil preparation may be blended and can constitute greater than about 10%, 25%, 35%, 50%, or 75% of the blend by volume. Oil produced from a plant of the present invention can be admixed with one or more organic solvents or petroleum distillates.

Plants of the present invention can be part of or generated from a breeding program. The choice of breeding method depends on the mode of plant reproduction, the heritability of the trait(s) being improved, and the type of cultivar used commercially (*e.g.*, F_1 hybrid cultivar, pureline cultivar, etc.). Selected, non-limiting approaches, for breeding the plants of the present invention are set forth below. A breeding program can be enhanced using marker assisted selection of the progeny of any cross. It is further understood that any commercial and non-commercial cultivars can be utilized in a breeding program. Factors such as, for example, emergence vigor, vegetative vigor, stress tolerance, disease resistance, branching, flowering, seed set, seed size, seed density, standability, and threshability etc. will generally dictate the choice.

For highly heritable traits, a choice of superior individual plants evaluated at a single location will be effective, whereas for traits with low heritability, selection should be based on mean values obtained from replicated evaluations of families of related plants. Popular selection methods commonly include pedigree selection, modified pedigree selection, mass selection, and recurrent selection. In a preferred embodiment a backcross or recurrent breeding program is undertaken.

The complexity of inheritance influences choice of the breeding method. Backcross breeding can be used to transfer one or a few favorable genes for a highly heritable trait into a desirable cultivar. This approach has been used extensively for breeding disease-resistant cultivars. Various recurrent selection techniques are used to improve quantitatively inherited traits controlled by numerous genes. The use of recurrent selection in self-pollinating crops depends on the ease of pollination, the frequency of successful hybrids from each pollination, and the number of hybrid offspring from each successful cross.

Breeding lines can be tested and compared to appropriate standards in environments representative of the commercial target area(s) for two or more generations. The best lines are candidates for new commercial cultivars; those still deficient in traits may be used as parents to produce new populations for further selection.

One method of identifying a superior plant is to observe its performance relative to other experimental plants and to a widely grown standard cultivar. If a single observation is inconclusive, replicated observations can provide a better estimate of its genetic worth. A breeder can select and cross two or more parental lines, followed by repeated selfing and selection, producing many new genetic combinations.

The development of new cultivars requires the development and selection of varieties, the crossing of these varieties and the selection of superior hybrid crosses. The hybrid seed can be produced by manual crosses between selected male-fertile parents or by using male sterility systems. Hybrids are selected for certain single gene traits such as pod color, flower color, seed yield, pubescence color, or herbicide resistance, which indicate that the seed is truly a hybrid. Additional data on parental lines, as well as the phenotype of the hybrid, influence the breeder's decision whether to continue with the specific hybrid cross.

Pedigree breeding and recurrent selection breeding methods can be used to develop cultivars from breeding populations. Breeding programs combine desirable traits from two or more cultivars or various broad-based sources into breeding pools from which cultivars are developed by selfing and selection of desired phenotypes. New cultivars can be evaluated to determine which have commercial potential.

Pedigree breeding is used commonly for the improvement of self-pollinating crops. Two parents who possess favorable, complementary traits are crossed to produce an F_1. A F_2 population is produced by selfing one or several F_1's. Selection of the best individuals from the best families is carried out. Replicated testing of families can begin in the F_4 generation to

improve the effectiveness of selection for traits with low heritability. At an advanced stage of inbreeding (*i.e.*, F_6 and F_7), the best lines or mixtures of phenotypically similar lines are tested for potential release as new cultivars.

Backcross breeding has been used to transfer genes for a simply inherited, highly heritable trait into a desirable homozygous cultivar or inbred line, which is the recurrent parent. The source of the trait to be transferred is called the donor parent. The resulting plant is expected to have the attributes of the recurrent parent (*e.g.*, cultivar) and the desirable trait transferred from the donor parent. After the initial cross, individuals possessing the phenotype of the donor parent are selected and repeatedly crossed (backcrossed) to the recurrent parent. The resulting parent is expected to have the attributes of the recurrent parent (*e.g.*, cultivar) and the desirable trait transferred from the donor parent.

The single-seed descent procedure in the strict sense refers to planting a segregating population, harvesting a sample of one seed per plant, and using the one-seed sample to plant the next generation. When the population has been advanced from the F_2 to the desired level of inbreeding, the plants from which lines are derived will each trace to different F_2 individuals. The number of plants in a population declines each generation due to failure of some seeds to germinate or some plants to produce at least one seed. As a result, not all of the F_2 plants originally sampled in the population will be represented by a progeny when generation advance is completed.

In a multiple-seed procedure, breeders commonly harvest one or more pods from each plant in a population and thresh them together to form a bulk. Part of the bulk is used to plant the next generation and part is put in reserve. The procedure has been referred to as modified single-seed descent or the pod-bulk technique.

The multiple-seed procedure has been used to save labor at harvest. It is considerably faster to thresh pods with a machine than to remove one seed from each by hand for the single-seed procedure. The multiple-seed procedure also makes it possible to plant the same number of seeds of a population each generation of inbreeding.

Descriptions of other breeding methods that are commonly used for different traits and crops can be found in one of several reference books (*e.g.*, Fehr, *Principles of Cultivar Development*, Vol. 1, pp. 2-3 (1987)).

A transgenic plant of the present invention may also be reproduced using apomixis. Apomixis is a genetically controlled method of reproduction in plants where the embryo is

formed without union of an egg and a sperm. There are three basic types of apomictic reproduction: 1) apospory where the embryo develops from a chromosomally unreduced egg in an embryo sac derived from the nucleus; 2) diplospory where the embryo develops from an unreduced egg in an embryo sac derived from the megaspore mother cell; and 3) adventitious embryony where the embryo develops directly from a somatic cell. In most forms of apomixis, pseudogamy, or fertilization of the polar nuclei to produce endosperm is necessary for seed viability. In apospory, a nurse cultivar can be used as a pollen source for endosperm formation in seeds. The nurse cultivar does not affect the genetics of the aposporous apomictic cultivar since the unreduced egg of the cultivar develops parthenogenetically, but makes possible endosperm production. Apomixis is economically important, especially in transgenic plants, because it causes any genotype, no matter how heterozygous, to breed true. Thus, with apomictic reproduction, heterozygous transgenic plants can maintain their genetic fidelity throughout repeated life cycles. Methods for the production of apomictic plants are known in the art. *See*, U.S. Patent 5,811,636.

OTHER ORGANISMS

A nucleic acid of the present invention may be introduced into any cell or organism such as a mammalian cell, mammal, fish cell, fish, bird cell, bird, algae cell, algae, fungal cell, fungi, or bacterial cell. A protein of the present invention may be produced in an appropriate cell or organism. Preferred host and transformants include: fungal cells such as *Aspergillus*, yeasts, mammals, particularly bovine and porcine, insects, bacteria, and algae. Particularly preferred bacteria are *Agrobacteruim tumefaciens* and *E. coli.*

Methods to transform such cells or organisms are known in the art (EP 0 238 023; Yelton *et al., Proc. Natl. Acad. Sci. (U.S.A.),* 81:1470-1474 (1984); Malardier *et al., Gene*, 78:147-156 (1989); Becker and Guarente, In: Abelson and Simon (eds.), *Guide to Yeast Genetics and Molecular Biology, Method Enzymol.,* Vol. 194, pp. 182-187, Academic Press, Inc., NY; Ito *et al., J. Bacteriology*, 153:163 (1983); Hinnen *et al., Proc. Natl. Acad. Sci. (U.S.A.),* 75:1920 (1978); Bennett and LaSure (eds.), *More Gene Manipualtionins in fungi*, Academic Press, CA (1991)). Methods to produce proteins of the present invention are also known (Kudla *et al., EMBO,* 9:1355-1364 (1990); Jarai and Buxton, *Current Genetics,* 26:2238-2244 (1994); Verdier, *Yeast,* 6:271-297 (1990); MacKenzie *et al., Journal of Gen. Microbiol.*, 139:2295-2307 (1993); Hartl *et al., TIBS,* 19:20-25 (1994); Bergenron *et al., TIBS,* 19:124-128 (1994); Demolder *et al., J. Biotechnology,* 32:179-189 (1994); Craig,

Science, 260:1902-1903 (1993); Gething and Sambrook, *Nature,* 355:33-45 (1992); Puig and Gilbert, J., *Biol. Chem.,* 269:7764-7771 (1994); Wang and Tsou, *FASEB Journal,* 7:1515-1517 (1993); Robinson *et al., Bio/Technology,* 1:381-384 (1994); Enderlin and Ogrydziak, *Yeast,* 10:67-79 (1994); Fuller *et al., Proc. Natl. Acad. Sci. (U.S.A.),* 86:1434-1438 (1989); Julius *et al., Cell,* 37:1075-1089 (1984); Julius *et al., Cell*, 32:839-852 (1983)).

In a preferred embodiment, overexpression of a protein or fragment thereof of the present invention in a cell or organism provides in that cell or organism, relative to an untransformed cell or organism with a similar genetic background, an increased level of tocopherols.

In a preferred embodiment, overexpression of a protein or fragment thereof of the present invention in a cell or organism provides in that cell or organism, relative to an untransformed cell or organism with a similar genetic background, an increased level of α-tocopherols.

In a preferred embodiment, overexpression of a protein or fragment thereof of the present invention in a cell or organism provides in that cell or organism, relative to an untransformed cell or organism with a similar genetic background, an increased level of γ-tocopherols.

In another preferred embodiment, overexpression of a protein or fragment thereof of the present invention in a cell or organism provides in that cell or organism, relative to an untransformed cell or organism with a similar genetic background, an increased level of α-tocotrienols.

In another preferred embodiment, overexpression of a protein or fragment thereof of the present invention in a cell or organism provides in that cell or organism, relative to an untransformed cell or organism with a similar genetic background, an increased level of γ-tocotrienols.

ANTIBODIES

One aspect of the present invention concerns antibodies, single-chain antigen binding molecules, or other proteins that specifically bind to one or more of the protein or peptide molecules of the present invention and their homologs, fusions or fragments. In a particularly preferred embodiment, the antibody specifically binds to a protein having the amino acid sequence set forth in SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, or fragments thereof.

Antibodies of the present invention may be used to quantitatively or qualitatively detect the protein or peptide molecules of the present invention, or to detect post translational modifications of the proteins. As used herein, an antibody or peptide is said to "specifically bind" to a protein or peptide molecule of the present invention if such binding is not competitively inhibited by the presence of non-related molecules.

Nucleic acid molecules that encode all or part of the protein of the present invention can be expressed, via recombinant means, to yield protein or peptides that can in turn be used to elicit antibodies that are capable of binding the expressed protein or peptide. Such antibodies may be used in immunoassays for that protein. Such protein-encoding molecules, or their fragments may be a "fusion" molecule (*i.e.*, a part of a larger nucleic acid molecule) such that, upon expression, a fusion protein is produced. It is understood that any of the nucleic acid molecules of the present invention may be expressed, via recombinant means, to yield proteins or peptides encoded by these nucleic acid molecules.

The antibodies that specifically bind proteins and protein fragments of the present invention may be polyclonal or monoclonal and may comprise intact immunoglobulins, or antigen binding portions of immunoglobulins fragments (such as (F(ab'), $F(ab')_2$)), or single-chain immunoglobulins producible, for example, via recombinant means. It is understood that practitioners are familiar with the standard resource materials that describe specific conditions and procedures for the construction, manipulation and isolation of antibodies (*see*, for example, Harlow and Lane, In: *Antibodies: A Laboratory Manual,* Cold Spring Harbor Press, Cold Spring Harbor, NY (1988)).

As discussed below, such antibody molecules or their fragments may be used for diagnostic purposes. Where the antibodies are intended for diagnostic purposes, it may be desirable to derivatize them, for example with a ligand group (such as biotin) or a detectable marker group (such as a fluorescent group, a radioisotope or an enzyme).

The ability to produce antibodies that bind the protein or peptide molecules of the present invention permits the identification of mimetic compounds derived from those molecules. These mimetic compounds may contain a fragment of the protein or peptide or merely a structurally similar region and nonetheless exhibits an ability to specifically bind to antibodies directed against that compound.

EXEMPLARY USES

Nucleic acid molecules and fragments thereof of the present invention may be employed to obtain other nucleic acid molecules from the same species (nucleic acid molecules from corn may be utilized to obtain other nucleic acid molecules from corn). Such nucleic acid molecules include the nucleic acid molecules that encode the complete coding sequence of a protein and promoters and flanking sequences of such molecules. In addition, such nucleic acid molecules include nucleic acid molecules that encode for other isozymes or gene family members. Such molecules can be readily obtained by using the above-described nucleic acid molecules or fragments thereof to screen cDNA or genomic libraries. Methods for forming such libraries are well known in the art.

Nucleic acid molecules and fragments thereof of the present invention may also be employed to obtain nucleic acid homologs. Such homologs include the nucleic acid molecules of plants and other organisms, including bacteria and fungi, including the nucleic acid molecules that encode, in whole or in part, protein homologues of other plant species or other organisms, sequences of genetic elements, such as promoters and transcriptional regulatory elements. Such molecules can be readily obtained by using the above-described nucleic acid molecules or fragments thereof to screen cDNA or genomic libraries obtained from such plant species. Methods for forming such libraries are well known in the art. Such homolog molecules may differ in their nucleotide sequences from those coding for one or more of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof because complete complementarity is not needed for stable hybridization. The nucleic acid molecules of the present invention therefore also include molecules that, although capable of specifically hybridizing with the nucleic acid molecules may lack "complete complementarity".

Any of a variety of methods may be used to obtain one or more of the above-described nucleic acid molecules (Zamechik *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 83:4143-4146 (1986); Goodchild *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 85:5507-5511 (1988); Wickstrom *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 85:1028-1032 (1988); Holt *et al.*, *Molec. Cell. Biol.*, 8:963-973 (1988); Gerwirtz *et al.*, *Science*, 242:1303-1306 (1988); Anfossi *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 86:3379-3383 (1989); Becker *et al.*, *EMBO J.*, 8:3685-3691 (1989)). Automated nucleic acid synthesizers may be employed for this purpose. In lieu of such synthesis, the disclosed nucleic acid molecules may be used to define a pair of primers that

can be used with the polymerase chain reaction (Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, 51:263-273 (1986); Erlich *et al.*, EP 50 424; EP 84 796; EP 258 017; EP 237 362; Mullis, EP 201 184; Mullis *et al.*, U.S. Patent 4,683,202; Erlich, U.S. Patent 4,582,788; and Saiki *et al.*, U.S. Patent 4,683,194) to amplify and obtain any desired nucleic acid molecule or fragment.

Promoter sequences and other genetic elements, including but not limited to transcriptional regulatory flanking sequences, associated with one or more of the disclosed nucleic acid sequences can also be obtained using the disclosed nucleic acid sequence provided herein. In one embodiment, such sequences are obtained by incubating nucleic acid molecules of the present invention with members of genomic libraries and recovering clones that hybridize to such nucleic acid molecules thereof. In a second embodiment, methods of "chromosome walking", or inverse PCR may be used to obtain such sequences (Frohman *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 85:8998-9002 (1988); Ohara *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 86:5673-5677 (1989); Pang *et al.*, *Biotechniques*, 22:1046-1048 (1977); Huang *et al.*, *Methods Mol. Biol.*, 69:89-96 (1997); Huang *et al.*, *Method Mol. Biol.*, 67:287-294 (1997); Benkel *et al.*, *Genet. Anal.*, 13:123-127 (1996); Hartl *et al.*, *Methods Mol. Biol.*, 58:293-301 (1996)). The term "chromosome walking" means a process of extending a genetic map by successive hybridization steps.

The nucleic acid molecules of the present invention may be used to isolate promoters of cell enhanced, cell specific, tissue enhanced, tissue specific, developmentally or environmentally regulated expression profiles. Isolation and functional analysis of the 5' flanking promoter sequences of these genes from genomic libraries, for example, using genomic screening methods and PCR techniques would result in the isolation of useful promoters and transcriptional regulatory elements. These methods are known to those of skill in the art and have been described (s*ee*, for example, Birren *et al.*, *Genome Analysis*: *Analyzing DNA*, 1, (1997), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Promoters obtained utilizing the nucleic acid molecules of the present invention could also be modified to affect their control characteristics. Examples of such modifications would include but are not limited to enhancer sequences. Such genetic elements could be used to enhance gene expression of new and existing traits for crop improvement.

Another subset of the nucleic acid molecules of the present invention includes nucleic acid molecules that are markers. The markers can be used in a number of conventional ways

in the field of molecular genetics. Such markers include nucleic acid molecules encoding SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90, and complements thereof, and fragments of either that can act as markers and other nucleic acid molecules of the present invention that can act as markers.

Genetic markers of the present invention include "dominant" or "codominant" markers. "Codominant markers" reveal the presence of two or more alleles (two per diploid individual) at a locus. "Dominant markers" reveal the presence of only a single allele per locus. The presence of the dominant marker phenotype (*e.g.*, a band of DNA) is an indication that one allele is in either the homozygous or heterozygous condition. The absence of the dominant marker phenotype (*e.g.*, absence of a DNA band) is merely evidence that "some other" undefined allele is present. In the case of populations where individuals are predominantly homozygous and loci are predominately dimorphic, dominant and codominant markers can be equally valuable. As populations become more heterozygous and multi-allelic, codominant markers often become more informative of the genotype than dominant markers. Marker molecules can be, for example, capable of detecting polymorphisms such as single nucleotide polymorphisms (SNPs).

The genomes of animals and plants naturally undergo spontaneous mutation in the course of their continuing evolution (Gusella, *Ann. Rev. Biochem.*, 55:831-854 (1986)). A "polymorphism" is a variation or difference in the sequence of the gene or its flanking regions that arises in some of the members of a species. The variant sequence and the "original" sequence co-exist in the species' population. In some instances, such co-existence is in stable or quasi-stable equilibrium.

A polymorphism is thus said to be "allelic", in that, due to the existence of the polymorphism, some members of a population may have the original sequence (*i.e.*, the original "allele") whereas other members may have the variant sequence (*i.e.*, the variant "allele"). In the simplest case, only one variant sequence may exist and the polymorphism is thus said to be di-allelic. In other cases, the species' population may contain multiple alleles and the polymorphism is termed tri-allelic, etc. A single gene may have multiple different unrelated polymorphisms. For example, it may have a di-allelic polymorphism at one site and a multi-allelic polymorphism at another site.

The variation that defines the polymorphism may range from a single nucleotide variation to the insertion or deletion of extended regions within a gene. In some cases, the

DNA sequence variations are in regions of the genome that are characterized by short tandem repeats (STRs) that include tandem di- or tri-nucleotide repeated motifs of nucleotides. Polymorphisms characterized by such tandem repeats are referred to as "variable number tandem repeat" ("VNTR") polymorphisms. VNTRs have been used in identity analysis (Weber, U.S. Patent 5,075,217; Armour *et al., FEBS Lett.*, 307:113-115 (1992); Jones *et al.*, *Eur. J. Haematol.*, 39:144-147 (1987); Horn *et al.*, PCT Application WO 91/14003; Jeffreys, EP 370 719; Jeffreys, U.S. Patent 5,175,082; Jeffreys *et al.*, *Amer. J. Hum. Genet.*, 39:11-24 (1986); Jeffreys *et al.*, *Nature*, 316:76-79 (1985); Gray *et al.*, *Proc. R. Acad. Soc. Lond.*, 243:241-253 (1991); Moore *et al.*, *Genomics*, 10:654-660 (1991); Jeffreys *et al.*, *Anim. Genet.*, 18:1-15 (1987); Hillel *et al.*, *Anim. Genet.*, 20:145-155 (1989); Hillel *et al.*, *Genet.*, 124:783-789 (1990)).

The detection of polymorphic sites in a sample of DNA may be facilitated through the use of nucleic acid amplification methods. Such methods specifically increase the concentration of polynucleotides that span the polymorphic site, or include that site and sequences located either distal or proximal to it. Such amplified molecules can be readily detected by gel electrophoresis or other means.

In an alternative embodiment, such polymorphisms can be detected through the use of a marker nucleic acid molecule that is physically linked to such polymorphism(s). For this purpose, marker nucleic acid molecules comprising a nucleotide sequence of a polynucleotide located within 1 mb of the polymorphism(s) and more preferably within 100kb of the polymorphism(s) and most preferably within 10kb of the polymorphism(s) can be employed.

The identification of a polymorphism can be determined in a variety of ways. By correlating the presence or absence of it in a plant with the presence or absence of a phenotype, it is possible to predict the phenotype of that plant. If a polymorphism creates or destroys a restriction endonuclease cleavage site, or if it results in the loss or insertion of DNA (*e.g.*, a VNTR polymorphism), it will alter the size or profile of the DNA fragments that are generated by digestion with that restriction endonuclease. As such, organisms that possess a variant sequence can be distinguished from those having the original sequence by restriction fragment analysis. Polymorphisms that can be identified in this manner are termed "restriction fragment length polymorphisms" (RFLPs) (Glassberg, UK Patent Application 2135774; Skolnick *et al.*, *Cytogen. Cell Genet.*, 32:58-67 (1982); Botstein *et al.*, *Ann. J. Hum.*

Genet., 32:314-331 (1980); Fischer *et al.*, PCT Application WO 90/13668; Uhlen, PCT Application WO 90/11369).

Polymorphisms can also be identified by Single Strand Conformation Polymorphism (SSCP) analysis (Elles, *Methods in Molecular Medicine: Molecular Diagnosis of Genetic Diseases*, Humana Press (1996)); Orita *et al., Genomics*, 5:874-879 (1989)). A number of protocols have been described for SSCP including, but not limited to, Lee *et al., Anal. Biochem.*, 205:289-293 (1992); Suzuki *et al., Anal. Biochem.*, 192:82-84 (1991); Lo *et al., Nucleic Acids Research*, 20:1005-1009 (1992); Sarkar *et al., Genomics*, 13:441-443 (1992). It is understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by SSCP analysis.

Polymorphisms may also be found using a DNA fingerprinting technique called amplified fragment length polymorphism (AFLP), which is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA to profile that DNA (Vos *et al.*, *Nucleic Acids Res.*, 23:4407-4414 (1995)). This method allows for the specific co-amplification of high numbers of restriction fragments, which can be visualized by PCR without knowledge of the nucleic acid sequence. It is understood that one or more of the nucleic acids of the present invention may be utilized as markers or probes to detect polymorphisms by AFLP analysis or for fingerprinting RNA.

Polymorphisms may also be found using random amplified polymorphic DNA (RAPD) (Williams *et al., Nucl. Acids Res.*, 18:6531-6535 (1990)) and cleaveable amplified polymorphic sequences (CAPS) (Lyamichev *et al.*, *Science*, 260:778-783 (1993)). It is understood that one or more of the nucleic acid molecules of the present invention, may be utilized as markers or probes to detect polymorphisms by RAPD or CAPS analysis.

Single Nucleotide Polymorphisms (SNPs) generally occur at greater frequency than other polymorphic markers and are spaced with a greater uniformity throughout a genome than other reported forms of polymorphism. The greater frequency and uniformity of SNPs means that there is greater probability that such a polymorphism will be found near or in a genetic locus of interest than would be the case for other polymorphisms. SNPs are located in protein-coding regions and noncoding regions of a genome. Some of these SNPs may result in defective or variant protein expression (*e.g.*, as a result of mutations or defective splicing). Analysis (genotyping) of characterized SNPs can require only a plus/minus assay rather than a lengthy measurement, permitting easier automation.

SNPs can be characterized using any of a variety of methods. Such methods include the direct or indirect sequencing of the site, the use of restriction enzymes (Botstein *et al.*, *Am. J. Hum. Genet.*, 32:314-331 (1980); Konieczny and Ausubel, *Plant J.*, 4:403-410 (1993)), enzymatic and chemical mismatch assays (Myers *et al.*, *Nature*, 313:495-498 (1985)), allele-specific PCR (Newton *et al.*, *Nucl. Acids Res.*, 17:2503-2516 (1989); Wu *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 86:2757-2760 (1989)), ligase chain reaction (Barany, *Proc. Natl. Acad. Sci. (U.S.A.)*, 88:189-193 (1991)), single-strand conformation polymorphism analysis (Labrune *et al.*, *Am. J. Hum. Genet.*, 48:1115-1120 (1991)), single base primer extension (Kuppuswamy *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 88:1143-1147 (1991), Goelet, U.S. Patent 6,004,744; Goelet, U.S. Patent 5,888,819), solid-phase ELISA-based oligonucleotide ligation assays (Nikiforov *et al.*, *Nucl. Acids Res.*, 22:4167-4175 (1994)), dideoxy fingerprinting (Sarkar *et al.*, *Genomics*, 13:441-443 (1992)), oligonucleotide fluorescence-quenching assays (Livak *et al.*, *PCR Methods Appl.*, 4:357-362 (1995a)), 5'-nuclease allele-specific hybridization TaqMan™ assay (Livak *et al.*, *Nature Genet.*, 9:341-342 (1995)), template-directed dye-terminator incorporation (TDI) assay (Chen and Kwok, *Nucl. Acids Res.*, 25:347-353 (1997)), allele-specific molecular beacon assay (Tyagi *et al.*, *Nature Biotech.*, 16:49-53 (1998)), PinPoint™ assay (Haff and Smirnov, *Genome Res.*, 7:378-388 (1997)), dCAPS analysis (Neff *et al.*, *Plant J.*, 14:387-392 (1998)), pyrosequencing (Ronaghi *et al.*, *Analytical Biochemistry*, 267:65-71 (1999); Ronaghi *et al.*, WO 98/13523; Nyren *et al.*, WO 98/28440) using mass spectrometry, *e.g.* the Masscode ™ system (Howbert *et al.*, WO 99/05319; Howbert *et al.*, WO 97/27331; Becker *et al.*, WO 98/26095; Becker *et al.*, WO 98/12355; Becker *et al.*, WO 97/33000; Monforte *et al.*, U.S. Patent 5,965,363), invasive cleavage of oligonucleotide probes (Lyamichev *et al.*, *Nature Biotechnology*, 17:292-296), and using high density oligonucleotide arrays (Hacia *et al.*, *Nature Genetics*, 22:164-167).

Polymorphisms may also be detected using allele-specific oligonucleotides (ASO), which, can be for example, used in combination with hybridization based technology including Southern, Northern, and dot blot hybridizations, reverse dot blot hybridizations and hybridizations performed on microarray and related technology.

The stringency of hybridization for polymorphism detection is highly dependent upon a variety of factors, including length of the allele-specific oligonucleotide, sequence composition, degree of complementarity (*i.e.*, presence or absence of base mismatches),

concentration of salts and other factors such as formamide and temperature. These factors are important both during the hybridization itself and during subsequent washes performed to remove target polynucleotide that is not specifically hybridized. In practice, the conditions of the final, most stringent wash are most critical. In addition, the amount of target polynucleotide that is able to hybridize to the allele-specific oligonucleotide is also governed by such factors as the concentration of both the ASO and the target polynucleotide, the presence and concentration of factors that act to "tie up" water molecules, so as to effectively concentrate the reagents (*e.g.*, PEG, dextran, dextran sulfate, *etc.*), whether the nucleic acids are immobilized or in solution, and the duration of hybridization and washing steps.

Hybridizations are preferably performed below the melting temperature (T_m) of the ASO. The closer the hybridization and/or washing step is to the T_m, the higher the stringency. T_m for an oligonucleotide may be approximated, for example, according to the following formula: $T_m = 81.5 + 16.6 \times (\log10[Na+]) + 0.41 \times (\%G+C) - 675/n$; where [Na+] is the molar salt concentration of Na+ or any other suitable cation and n = number of bases in the oligonucleotide. Other formulas for approximating T_m are available and are known to those of ordinary skill in the art.

Stringency is preferably adjusted so as to allow a given ASO to differentially hybridize to a target polynucleotide of the correct allele and a target polynucleotide of the incorrect allele. Preferably, there will be at least a two-fold differential between the signal produced by the ASO hybridizing to a target polynucleotide of the correct allele and the level of the signal produced by the ASO cross-hybridizing to a target polynucleotide of the incorrect allele (*e.g.*, an ASO specific for a mutant allele cross-hybridizing to a wild-type allele). In more preferred embodiments of the present invention, there is at least a five-fold signal differential. In highly preferred embodiments of the present invention, there is at least an order of magnitude signal differential between the ASO hybridizing to a target polynucleotide of the correct allele and the level of the signal produced by the ASO cross-hybridizing to a target polynucleotide of the incorrect allele.

While certain methods for detecting polymorphisms are described herein, other detection methodologies may be utilized. For example, additional methodologies are known and set forth, in Birren *et al.*, *Genome Analysis,* 4:135-186; *A Laboratory Manual. Mapping Genomes*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1999); Maliga *et al.*, *Methods in Plant Molecular Biology. A Laboratory Course Manual*, Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, NY (1995); Paterson, *Biotechnology Intelligence Unit: Genome Mapping in Plants*, R.G. Landes Co., Georgetown, TX, and Academic Press, San Diego, CA (1996); *The Corn Handbook*, Freeling and Walbot, (eds.), Springer-Verlag, New York, NY (1994); *Methods in Molecular Medicine: Molecular Diagnosis of Genetic Diseases*, Elles, (ed.), Humana Press, Totowa, NJ (1996); Clark, (ed.), *Plant Molecular Biology: A Laboratory Manual*, Springer-Verlag, Berlin, Germany (1997).

Factors for marker-assisted selection in a plant breeding program are: (1) the marker(s) should co-segregate or be closely linked with the desired trait; (2) an efficient means of screening large populations for the molecular marker(s) should be available; and (3) the screening technique should have high reproducibility across laboratories and preferably be economical to use and be user-friendly.

The genetic linkage of marker molecules can be established by a gene mapping model such as, without limitation, the flanking marker model reported by Lander and Botstein, *Genetics*, 121:185-199 (1989) and the interval mapping, based on maximum likelihood methods described by Lander and Botstein, *Genetics*, 121:185-199 (1989) and implemented in the software package MAPMAKER/QTL (Lincoln and Lander, *Mapping Genes Controlling Quantitative Traits Using MAPMAKER/QTL,* Whitehead Institute for Biomedical Research, MA (1990). Additional software includes Qgene, Version 2.23 (1996), Department of Plant Breeding and Biometry, 266 Emerson Hall, Cornell University, Ithaca, NY. Use of Qgene software is a particularly preferred approach.

A maximum likelihood estimate (MLE) for the presence of a marker is calculated, together with an MLE assuming no QTL effect, to avoid false positives. A $\log_{10}$ of an odds ratio (LOD) is then calculated as: LOD = $\log_{10}$ (MLE for the presence of a QTL/MLE given no linked QTL).

The LOD score essentially indicates how much more likely the data are to have arisen assuming the presence of a QTL than in its absence. The LOD threshold value for avoiding a false positive with a given confidence, say 95%, depends on the number of markers and the length of the genome. Graphs indicating LOD thresholds are set forth in Lander and Botstein, *Genetics*, 121:185-199 (1989) and further described by Arús and Moreno-González, *Plant Breeding*, Hayward *et al.,* (eds.) Chapman & Hall, London, pp. 314-331 (1993).

In a preferred embodiment of the present invention the nucleic acid marker exhibits a LOD score of greater than 2.0, more preferably 2.5, even more preferably greater than 3.0 or

4.0 with the trait or phenotype of interest. In a preferred embodiment, the trait of interest is altered tocopherol levels or compositions or altered tocotrienol levels or compositions.

Additional models can be used. Many modifications and alternative approaches to interval mapping have been reported, including the use of non-parametric methods (Kruglyak and Lander, *Genetics*, 139:1421-1428 (1995)). Multiple regression methods or models can also be used, in which the trait is regressed on a large number of markers (Jansen, *Biometrics in Plant Breeding*, van Oijen and Jansen (eds.), Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, pp. 116-124 (1994); Weber and Wricke, *Advances in Plant Breeding,* Blackwell, Berlin, 16 (1994)). Procedures combining interval mapping with regression analysis, whereby the phenotype is regressed onto a single putative QTL at a given marker interval and at the same time onto a number of markers that serve as "cofactors", have been reported by Jansen and Stam, *Genetics*, 136:1447-1455 (1994); and Zeng, *Genetics*, 136:1457-1468 (1994). Generally, the use of cofactors reduces the bias and sampling error of the estimated QTL positions (Utz and Melchinger, *Biometrics in Plant Breeding,* van Oijen and Jansen (eds.), Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, pp.195-204 (1994), thereby improving the precision and efficiency of QTL mapping (Zeng, *Genetics*, 136:1457-1468 (1994)). These models can be extended to multi-environment experiments to analyze genotype-environment interactions (Jansen *et al., Theo. Appl. Genet.*, 91:33-37 (1995)).

It is understood that one or more of the nucleic acid molecules of the present invention may be used as molecular markers. It is also understood that one or more of the protein molecules of the present invention may be used as molecular markers.

In a preferred embodiment, the polymorphism is present and screened for in a mapping population, *e.g.* a collection of plants capable of being used with markers such as polymorphic markers to map genetic position of traits. The choice of appropriate mapping population often depends on the type of marker systems employed (Tanksley *et al.,* J.P. Gustafson and R. Appels (eds.). Plenum Press, NY, pp. 157-173 (1988)). Consideration must be given to the source of parents (adapted vs. exotic) used in the mapping population. Chromosome pairing and recombination rates can be severely disturbed (suppressed) in wide crosses (adapted x exotic) and generally yield greatly reduced linkage distances. Wide

crosses will usually provide segregating populations with a relatively large number of polymorphisms when compared to progeny in a narrow cross (adapted x adapted).

An F_2 population is the first generation of selfing (self-pollinating) after the hybrid seed is produced. Usually a single F_1 plant is selfed to generate a population segregating for all the genes in Mendelian (1:2:1) pattern. Maximum genetic information is obtained from a completely classified F_2 population using a codominant marker system (Mather, Measurement of Linkage in Heredity: Methuen and Co., (1938)). In the case of dominant markers, progeny tests (*e.g.*, F_3, BCF_2) are required to identify the heterozygotes, in order to classify the population. However, this procedure is often prohibitive because of the cost and time involved in progeny testing. Progeny testing of F_2 individuals is often used in map construction where phenotypes do not consistently reflect genotype (*e.g.*, disease resistance) or where trait expression is controlled by a QTL. Segregation data from progeny test populations (*e.g.*, F_3 or BCF_2) can be used in map construction. Marker-assisted selection can then be applied to cross progeny based on marker-trait map associations (F_2, F_3), where linkage groups have not been completely disassociated by recombination events (*i.e.*, maximum disequilibrium).

Recombinant inbred lines (RIL) (genetically related lines; usually >F_5, developed from continuously selfing F_2 lines towards homozygosity) can be used as a mapping population. Information obtained from dominant markers can be maximized by using RIL because all loci are homozygous or nearly so. Under conditions of tight linkage (*i.e.*, about <10% recombination), dominant and co-dominant markers evaluated in RIL populations provide more information per individual than either marker type in backcross populations (Reiter *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 89:1477-1481 (1992)). However, as the distance between markers becomes larger (*i.e.*, loci become more independent), the information in RIL populations decreases dramatically when compared to codominant markers.

Backcross populations *e.g.*, generated from a cross between a successful variety (recurrent parent) and another variety (donor parent) carrying a trait not present in the former) can be utilized as a mapping population. A series of backcrosses to the recurrent parent can be made to recover most of its desirable traits. Thus a population is created consisting of individuals nearly like the recurrent parent but each individual carries varying amounts or mosaic of genomic regions from the donor parent. Backcross populations can be useful for mapping dominant markers if all loci in the recurrent parent are homozygous and the donor

and recurrent parent have contrasting polymorphic marker alleles (Reiter *et al., Proc. Natl. Acad. Sci. (U.S.A.)*, 89:1477-1481 (1992)). Information obtained from backcross populations using either codominant or dominant markers is less than that obtained from F_2 populations because one, rather than two, recombinant gamete is sampled per plant. Backcross populations, however, are more informative (at low marker saturation) when compared to RILs as the distance between linked loci increases in RIL populations (*i.e.*, about 0.15% recombination). Increased recombination can be beneficial for resolution of tight linkages, but may be undesirable in the construction of maps with low marker saturation.

Near-isogenic lines (NIL) (created by many backcrosses to produce a collection of individuals that is nearly identical in genetic composition except for the trait or genomic region under interrogation) can be used as a mapping population. In mapping with NILs, only a portion of the polymorphic loci is expected to map to a selected region.

Bulk segregant analysis (BSA) is a method developed for the rapid identification of linkage between markers and traits of interest (Michelmore *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)*, 88:9828-9832 (1991)). In BSA, two bulked DNA samples are drawn from a segregating population originating from a single cross. These bulks contain individuals that are identical for a particular trait (resistant or susceptible to particular disease) or genomic region but arbitrary at unlinked regions (*i.e.*, heterozygous). Regions unlinked to the target region will not differ between the bulked samples of many individuals in BSA.

In an aspect of the present invention, one or more of the nucleic molecules of the present invention are used to determine the level (*i.e.*, the concentration of mRNA in a sample, *etc.*) in a plant (preferably canola, corn, *Brassica campestris,* oilseed rape, rapeseed, soybean, crambe, mustard, castor bean, peanut, sesame, cottonseed, linseed, safflower, oil palm, flax or sunflower) or pattern (*i.e.*, the kinetics of expression, rate of decomposition, stability profile, *etc.*) of the expression of a protein encoded in part or whole by one or more of the nucleic acid molecule of the present invention (collectively, the "Expression Response" of a cell or tissue).

As used herein, the Expression Response manifested by a cell or tissue is said to be "altered" if it differs from the Expression Response of cells or tissues of plants not exhibiting the phenotype. To determine whether a Expression Response is altered, the Expression Response manifested by the cell or tissue of the plant exhibiting the phenotype is compared with that of a similar cell or tissue sample of a plant not exhibiting the phenotype. As will be

appreciated, it is not necessary to re-determine the Expression Response of the cell or tissue sample of plants not exhibiting the phenotype each time such a comparison is made; rather, the Expression Response of a particular plant may be compared with previously obtained values of normal plants. As used herein, the phenotype of the organism is any of one or more characteristics of an organism (*e.g.*, disease resistance, pest tolerance, environmental tolerance such as tolerance to abiotic stress, male sterility, quality improvement or yield *etc*.). A change in genotype or phenotype may be transient or permanent. Also as used herein, a tissue sample is any sample that comprises more than one cell. In a preferred aspect, a tissue sample comprises cells that share a common characteristic (*e.g.*, Derived from root, seed, flower, leaf, stem or pollen etc.).

In one aspect of the present invention, an evaluation can be conducted to determine whether a particular mRNA molecule is present. One or more of the nucleic acid molecules of the present invention are utilized to detect the presence or quantity of the mRNA species. Such molecules are then incubated with cell or tissue extracts of a plant under conditions sufficient to permit nucleic acid hybridization. The detection of double-stranded probe-mRNA hybrid molecules is indicative of the presence of the mRNA; the amount of such hybrid formed is proportional to the amount of mRNA. Thus, such probes may be used to ascertain the level and extent of the mRNA production in a plant's cells or tissues. Such nucleic acid hybridization may be conducted under quantitative conditions (thereby providing a numerical value of the amount of the mRNA present). Alternatively, the assay may be conducted as a qualitative assay that indicates either that the mRNA is present, or that its level exceeds a user set, predefined value.

A number of methods can be used to compare the expression response between two or more samples of cells or tissue. These methods include hybridization assays, such as northerns, RNAse protection assays, and *in situ* hybridization. Alternatively, the methods include PCR-type assays. In a preferred method, the expression response is compared by hybridizing nucleic acids from the two or more samples to an array of nucleic acids. The array contains a plurality of suspected sequences known or suspected of being present in the cells or tissue of the samples.

An advantage of *in situ* hybridization over more conventional techniques for the detection of nucleic acids is that it allows an investigator to determine the precise spatial population (Angerer *et al., Dev. Biol.*, 101:477-484 (1984); Angerer *et al., Dev. Biol.*,

112:157-166 (1985); Dixon *et al.*, *EMBO J.*, 10:1317-1324 (1991)). *In situ* hybridization may be used to measure the steady-state level of RNA accumulation (Hardin *et al.*, *J. Mol. Biol.*, 202:417-431 (1989)). A number of protocols have been devised for *in situ* hybridization, each with tissue preparation, hybridization and washing conditions (Meyerowitz, *Plant Mol. Biol. Rep.*, 5:242-250 (1987); Cox and Goldberg, In: *Plant Molecular Biology: A Practical Approach,* Shaw (ed.), pp. 1-35, IRL Press, Oxford (1988); Raikhel *et al.*, *In situ RNA hybridization in plant tissues*, In: *Plant Molecular Biology Manual*, Vol. B9:1-32, Kluwer Academic Publisher, Dordrecht, Belgium (1989)).

In situ hybridization also allows for the localization of proteins within a tissue or cell (Wilkinson, *In Situ Hybridization,* Oxford University Press, Oxford (1992); Langdale, *In Situ Hybridization* In: *The Corn Handbook,* Freeling and Walbot (eds.), pp. 165-179, Springer-Verlag, NY (1994)). It is understood that one or more of the molecules of the present invention, preferably one or more of the nucleic acid molecules or fragments thereof of the present invention or one or more of the antibodies of the present invention may be utilized to detect the level or pattern of a protein or mRNA thereof by *in situ* hybridization.

Fluorescent *in situ* hybridization allows the localization of a particular DNA sequence along a chromosome, which is useful, among other uses, for gene mapping, following chromosomes in hybrid lines, or detecting chromosomes with translocations, transversions or deletions. *In situ* hybridization has been used to identify chromosomes in several plant species (Griffor *et al., Plant Mol. Biol.*, 17:101-109 (1991); Gustafson *et al., Proc. Natl. Acad. Sci. (U.S.A.)*, 87:1899-1902 (1990); Mukai and Gill, *Genome*, 34:448-452 (1991); Schwarzacher and Heslop-Harrison, *Genome*, 34:317-323 (1991); Wang *et al., Jpn. J. Genet.*, 66:313-316 (1991); Parra and Windle, *Nature Genetics*, 5:17-21 (1993)). It is understood that the nucleic acid molecules of the present invention may be used as probes or markers to localize sequences along a chromosome.

Another method to localize the expression of a molecule is tissue printing. Tissue printing provides a way to screen, at the same time on the same membrane many tissue sections from different plants or different developmental stages (Yomo and Taylor, *Planta*, 112:35-43 (1973); Harris and Chrispeels, *Plant Physiol.*, 56:292-299 (1975); Cassab and Varner, *J. Cell. Biol.*, 105:2581-2588 (1987); Spruce *et al.*, *Phytochemistry*, 26:2901-2903 (1987); Barres *et al., Neuron*, 5:527-544 (1990); Reid and Pont-Lezica, *Tissue Printing: Tools for the Study of Anatomy, Histochemistry and Gene Expression,* Academic Press, New York,

NY (1992); Reid *et al., Plant Physiol.*, 93:160-165 (1990); Ye *et al.*, *Plant J.*, 1:175-183 (1991)).

One skilled in the art can refer to general reference texts for detailed descriptions of known techniques discussed herein or equivalent techniques. These texts include *Current Protocols in Molecular Biology*, Ausubel *et al.,* (eds.), John Wiley & Sons, NY (1989), and supplements through September (1998), *Molecular Cloning, A Laboratory Manual*, Sambrook *et al.,* 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), *Genome Analysis: A Laboratory Manual* 1: *Analyzing DNA*, Birren *et al.,* Cold Spring Harbor Press, Cold Spring Harbor, NY (1997); *Genome Analysis: A Laboratory Manual* 2: *Detecting Genes*, Birren *et al.,* Cold Spring Harbor Press, Cold Spring Harbor, NY (1998); *Genome Analysis: A Laboratory Manual* 3: *Cloning Systems*, Birren *et al.,* Cold Spring Harbor Press, Cold Spring Harbor, NY (1999); *Genome Analysis: A Laboratory Manual* 4: *Mapping Genomes*, Birren *et al.,* Cold Spring Harbor Press, Cold Spring Harbor, NY (1999); *Plant Molecular Biology: A Laboratory Manual*, Clark, Springer-Verlag, Berlin, (1997); *Methods in Plant Molecular Biology*, Maliga *et al.,* Cold Spring Harbor Press, Cold Spring Harbor, NY (1995). These texts can, of course, also be referred to in making or using an aspect of the present invention. It is understood that any of the agents of the present invention can be substantially purified and/or be biologically active and/or recombinant.

Having now generally described the present invention, the same will be more readily understood through reference to the following examples that are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

EXAMPLE 1: Identification of homogentisate prenyl transferase Sequences

This example sets forth methods used to analyze homogentisate prenyl transferase sequences from various sources in order to identify motifs common to homogentisate prenyl transferase that are contained therein.

Homogentisate prenyl transferase sequences from Soy, *Arabidopsis*, Corn and *Cuphea* (partial) are cloned and sequenced from EST sequences found in an EST database. *Synechocystis*, *Nostoc*, and *Anabaena* are obtained from Genbank. These sequences (representing SEQ ID NOs: 1-8) are then aligned with respect to each other using the multiple alignment software ClustalX, which is described by Thompson *et al.*, *Nucleic Acids Research*, 24:4876-4882 (1997). The multiple alignment of the protein sequences is visualized and edited using Genedoc, which is described by Nicholas *et al.*, EMBNEW.NEWS, 4:14 (1997).

Using the aforementioned multiple alignment tool, four motifs (A-D) are identified, as shown in Figures 2a-2c, wherein motifs A-D are set forth. These motifs are represented by SEQ ID NOs: 12-15. The *Cuphea* sequence is removed from motif D because the sequence had multiple errors towards the 3' end that generated apparent frame shift errors.

The specificity of these motifs is demonstrated using a Hidden Markov Model (HMM) that is built using an HMMER(version 2.2g) software package (Eddy, *Bioinformatics,* 14:755-763 (1998)). A HMM search is performed on a cDNA sequence database containing full insert sequence from different plant species. This search identifies two new homogentisate prenyl transferase sequences (SEQ ID NOs: 9-10) in addition to several partial homogentisate prenyl transferase sequences. The two new homogentisate prenyl transferase sequences identified are from leek and wheat. This search also identifies a complete *Cuphea* sequence (SEQ ID NO: 11) with no errors. A second alignment is generated using the aforementioned multiple alignment tool, as shown in Figures 3a-3c. This alignment has the leek, wheat, and full *Cuphea* sequences incorporated. Motifs I-IV (SEQ ID NOs: 39-42) are shown.

Specificity is also tested by using each motif sequence to search the non-redundant amino acid database downloaded from Genbank available through NCBI. All four motifs identify three homogentisate prenyl transferase found in the aforementioned non-redundant amino acid database, as follows: *Nostoc*, *Synechocystis*, *Arabidopsis*. Motifs II and IV also identified some genomic variants of an uncharacterized *Arabidopsis* protein. Motifs I and III only identified known homogentisate prenyl transferase at an E value of 0.001 or lower.

EXAMPLE 2: Preparation of Expression Constructs

A plasmid containing the napin cassette derived from pCGN3223 (described in U.S. Patent 5,639,790) is modified to
make it more useful for cloning large DNA fragments containing multiple restriction sites, and to allow the cloning of multiple napin fusion genes into plant binary transformation vectors. An adapter comprised of the self annealed oligonucleotide of sequence
CGCGATTTAAATGGCGCGCCCTGCAGGCGGCCGCCTGCAGGGCGCGCCATTTAAA T (SEQ ID NO: 16) is ligated into the cloning vector pBC SK+ (Stratagene) after digestion with the restriction endonuclease BssHII to construct vector pCGN7765. Plasmids pCGN3223 and pCGN7765 are digested with NotI and ligated together. The resultant vector,

pCGN7770, contains the pCGN7765 backbone with the napin seed specific expression cassette from pCGN3223.

The cloning cassette pCGN7787 comprises essentially the same regulatory elements as pCGN7770, with the exception that the napin regulatory regions of pCGN7770 have been replaced with the double CAMV 35S promoter and the polyadenylation and transcriptional termination region.

A binary vector for plant transformation, pCGN5139, is constructed from pCGN1558 (McBride and Summerfelt, *Plant Molecular Biology*, 14:269-276 (1990)). The polylinker of pCGN1558 is replaced as a HindIII/Asp718 fragment with a polylinker containing unique restriction endonuclease sites, AscI, PacI, XbaI, SwaI, BamHI, and NotI. The Asp718 and HindIII restriction endonuclease sites are retained in pCGN5139.

A series of binary vectors are constructed to allow for the rapid cloning of DNA sequences into binary vectors containing transcriptional initiation regions (promoters) and transcriptional termination regions.

The plasmid pCGN8618 is constructed by ligating oligonucleotides 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO: 17) and 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO: 18) into SalII/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker and napin 3' region is excised from pCGN8618 by digestion with Asp718I; the fragment is blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that is digested with Asp7181 and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter is closest to the blunted Asp718I site of pCGN5139 and the napin 3' is closest to the blunted HindIII site is subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid is designated pCGN8622.

The plasmid pCGN8619 is constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO: 19) and 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO: 20) into SalI/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker and napin 3' region is removed from pCGN8619 by digestion with Asp718I; the fragment is blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that is digested with Asp7181 and HindIII and blunt-ended by filling in the 5' overhangs with

Klenow fragment. A plasmid containing the insert oriented so that the napin promoter is closest to the blunted Asp7181 site of pCGN5139 and the napin 3' is closest to the blunted HindIII site is subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid is designated pCGN8623.

The plasmid pCGNS620 is constructed by ligating oligonucleotides 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGGAGCT-3' (SEQ ID NO: 21) and 5'-CCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO: 22) into SalI/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and trial 3' region is removed from pCGN8620 by complete digestion with Asp718I and partial digestion with NotI. The fragment is blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that is digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter is closest to the blunted Asp718I site of pCGN5139 and the tml 3' is closest to the blunted HindIII site is subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid is designated pCGN8624.

The plasmid pCGN8621 is constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCCAGCT-3' (SEQ ID NO: 23) and 5'-GGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO: 24) into SalI/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region is removed from pCGN8621 by complete digestion with Asp718I and partial digestion with NotI.

The fragment is blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter is closest to the blunted Asp718I site of pCGN5139 and the tml 3' is closest to the blunted HindIII site is subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid is designated pCGN8625.

The plasmid construct pCGN8640 is a modification of pCGN8624 described above. A 938bp PstI fragment isolated from transposon Tn7 which encodes bacterial spectinomycin and streptomycin resistance (Fling *et al.*, *Nucleic Acids Research*, 13(19):7095-7106 (1985)), a

determinant for *E. coli* and Agrobacterium selection, is blunt-ended with Pfu polymerase. The blunt-ended fragment is ligated into pCGN8624 that had been digested with SpeI and blunt-ended with Pfu polymerase. The region containing the PstI fragment is sequenced to confirm both the insert orientation and the integrity of cloning junctions.

The spectinomycin resistance marker is introduced into pCGN8622 and pCGN8623 as follows. A 7.7 Kbp AvrII-SnaBI fragment from pCGN8640 is ligated to a 10.9 Kbp AvrII-SnaBI fragment from pCGN8623 or pCGN8622, described above. The resulting plasmids are pCGN8641 and pCGN8643, respectively.

The plasmid pCGN8644 is constructed by ligating oligonucleotides 5'-GATCACCTGCAGGAAGCTTGCGGCCGCGGATCCAATGCA-3' (SEQ ID NO: 25) and 5' TTGGATCCGCGGCCGCAAGCTTCCTGCAGGT-3' (SEQ ID NO: 26) into BamHI-PstI digested pCGN8640.

Synthetic oligonucleotides are designed for use in Polymerase Chain Reactions (PCR) to amplify the coding sequences of each of the nucleic acids that encode the polypeptides of SEQ ID NOs: 1-7, 9-11, 43-44, 57-58, and 90 for the preparation of expression constructs.

The coding sequences of each of the nucleic acids that encode the polypeptides of SEQ ID NOs: 1-7, 9-11, 43-44, 57-58, and 90 are all amplified and cloned into the TopoTA vector (Invitrogen). Constructs containing the respective homogentisate prenyl transferase sequences are digested with NotI and Sse8387I and cloned into the turbobinary vectors described above.

Synthetic oligonucleotides were designed for use in Polymerase Chain Reactions (PCR) to amplify SEQ ID NO: 33 for the preparation of expression constructs and are provided in the table below:

Restriction Site	Sequence	SEQ ID NO:
5' NotI	GGATCCGCGGCCGCACAATGG AGTCTCTGCTCTCTAGTTCT	37
3' SseI	GGATCCTGCAGGTCACTTCAAA AAAGGTAACAGCAAGT	38

SEQ ID NO: 33 was amplified using the respective PCR primers shown in the table above and cloned into the TopoTA vector (Invitrogen). Constructs containing the respective homogentisate prenyl transferase sequences were digested with NotI and Sse8387I and cloned into the turbobinary vectors described above.

SEQ ID NO: 33 was cloned in the sense orientation into pCGN8640 to produce the plant transformation construct pCGN10800 (Figure 4). SEQ ID NO: 33 is under control of the enhanced 35S promoter.

SEQ ID NO: 33 was also cloned in the antisense orientation into the construct pCGN8641 to create pCGN10801 (Figure 5). This construct provides for the antisense expression of SEQ ID NO: 33 from the napin promoter.

SEQ ID NO: 33 was also cloned in the sense orientation into the vector pCGN8643 to create the plant transformation construct pCGN10822 (Figure 7). This construct provides for the sense expression of SEQ ID NO: 33 from the napin promoter.

SEQ ID NO: 33 was also cloned in the antisense orientation into the vector pCGN8644 to create the plant transformation construct pCGN10803 (Figure 6). This construct provides for the antisense expression of SEQ ID NO: 33 from the enhanced 35S promoter.

EXAMPLE 3: Plant Transformation

Transgenic *Brassica* plants are obtained by *Agrobacte*rium-mediated transformation as described by Radke *et al., Theor. Appl. Genet.*, 75:685-694 (1988); *Plant Cell Reports*, 11:499-505 (1992). Transgenic *Arabidopsis thaliana* plants may be obtained *by Agrobacterium*-mediated transformation as described by Valverkens *et al., Proc. Nat. Acad. Sci.*, 85:5536-5540 (1988), or as described by Bent *et al.*, *Science*, 265:1856-1860 (1994), or Bechtold *et al.*, *C.R.Acad.Sci. Life Sciences*, 316:1194-1199 (1993). Other plant species may be similarly transformed using related techniques.

Alternatively, microprojectile bombardment methods, such as described by Klein *et al., Bio/Technology*, 10:286-291 may also be used to obtain nuclear transformed plants.

EXAMPLE 4: Identification of additional homogentisate prenyl transferase

In order to identify additional homogentisate prenyl transferase, motifs identified through sequence homology are used to search a database of cDNA sequences containing full insert sequences. The cDNA database is first translated in all six frames and then a HMM search is done using a HMM model built for the motifs. All HMM hits are annotated by performing a blast search against a non-redundant amino acid database. All motifs are sensitive and identify homogentisate prenyl transferase sequences present in the database. Novel homogentisate prenyl transferase sequences are thereby discovered.

EXAMPLE 5: Transgenic Plant Analysis

Arabidopsis plants transformed with constructs for the sense or antisense expression of the homogentisate prenyl transferase proteins are analyzed by High Performance Liquid Chromatography (HPLC) for altered levels of total tocopherols and tocotrienols, as well as altered levels of specific tocopherols and tocotrienols (e.g. α, β, γ, and δ-tocopherol/tocotrienol).

Extracts of leaves and seeds are prepared for HPLC as follows. For seed extracts, 10 mg of seed is added to 1 g of microbeads (Biospec) in a sterile microfuge tube to which 500 ul 1% pyrogallol (Sigma Chem)/ethanol is added. The mixture is shaken for 3 minutes in a mini Beadbeater (Biospec) on "fast" speed. The extract is filtered through a 0.2 um filter into an autosampler tube. The filtered extracts are then used in HPLC analysis described below.

Leaf extracts are prepared by mixing 30-50 mg of leaf tissue with 1 g microbeads and freezing in liquid nitrogen until extraction. For extraction, 500 ul 1% pyrogallol in ethanol is added to the leaf/bead mixture and shaken for 1 minute on a Beadbeater (Biospec) on "fast" speed. The resulting mixture is centrifuged for 4 minutes at 14,000 rpm and filtered as described above prior to HPLC analysis.

HPLC is performed on a Zorbax silica HPLC column (4.6 mm X 250 mm), using a fluorescent detection monitor, with excitation and emission spectra set at 290 nm and 336 nm, respectively. Solvent A is hexane and solvent B is methyl-t-butyl ether. The injection volume is 20 ul, the flow rate is 1.5 ml/min, the run time is 12 min (40°C) using the table below:

Time	Solvent A	Solvent B
0 min.	90%	10%
10 min.	90%	10%
11 min.	25%	75%
12 min.	90%	10%

Tocopherol standards in 1% pyrogallol/ethanol are also run for comparison (alpha tocopherol, gamma tocopherol, beta tocopherol, delta tocopherol, and tocopherol (tocol) (all from Matreya, State College, PA, or Calbiochem, La Jolla, CA.)).

Standard curves for alpha, beta, delta, and gamma tocopherol are calculated using Chemstation software. The absolute amount of component x is: Absolute amount of x= $Response_x$ x RF_x x dilution factor where $Response_x$ is the area of peak x, RF_x, is the response

factor for component x ($Amount_x/Response_x$) and the dilution factor is 500 ul. The ng/mg tissue is found by: total ng component/mg plant tissue.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing pMON10822 for the expression of SEQ ID NO: 33 from the napin promoter are provided in Figure 8.

HPLC analysis results of *Arabidopsis* seed tissue expressing the SEQ ID NO: 33 sequence from the napin promoter (pMON10822) demonstrates an increased level of tocopherols in the seed. Total tocopherol levels are increased as much as 50 to 60% over the total tocopherol levels of non-transformed (wild-type) *Arabidopsis* plants (Figure 8).

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines 1387-1624 containing pMON10803 for the antisense expression of SEQ ID NO: 33 from the enhanced 35S promoter are provided in Figure 9. Two lines, 1393 and 1401, show a substantial reduction in overall tocopherol levels, supporting the position that *HPT* is a homogentisate prenyl transferase involved in the synthesis of tocopherol.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing constructs for the expression of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90 are obtained.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing constructsfor the expression of SEQ ID NOs: 5, 9-11, 43-44, 57-58, and 90 from the enhanced 35S promoter are obtained.

EXAMPLE 6: Expression of a homogentisate prenyl transferase as single gene, and in combination with HPPD and *tyrA* in soy

The *Arabidopsis* homogentisate prenyl transferase (ATPT2) (SEQ ID NO: 33) was cloned in a soy binary vector harboring an Arcelin 5 expression cassette. This expression cassette consisted of an Arcelin 5-promoter, a multi cloning site, and the Arcelin 5 3'-untranslated sequence in the order as described. Vector construction for this construct and the following constructs was performed using standard cloning techniques well established in the art and described in lab manuals such as (Sambrook *et al.* 2001). The resulting binary vector for soy seed-specific expression of ATPT2 was designated pMON36581 (Figure 10). Similarly the *Synechocystis* homogentisate prenyl transferase (*slr1736*) (SEQ ID NO: 29) was fused to a chloroplast target peptide (CTP1), and cloned into the Arcelin 5 soy seed-specific expression cassette. The resulting binary plasmid was designated pMON69933 (Figure 11).

An additional binary plasmid for seed-specific co-expression of the *Arabidopsis* *p*-hydroxyphenylpyruvate dioxygenase ($HPPD_{At}$) and the bifunctional prephenate dehydrogenase from *Erwinia herbicola* (*tyrA*$_{Eh}$) (*see* WO 02/089561) was constructed by fusing the $HPPD_{At}$-gene and the *tyrA*$_{Eh}$-gene to the chloroplast target peptides, CTP2, and CTP1, respectively. These fusion genes were subsequently cloned into the multi cloning site of soy seed-specific expression cassettes consisting of the p7Sα'-promoter, a multi cloning site, and the E9 3'-untranslated region. The $HPPD_{At}$ expression cassette was cloned into a binary vector downstream of the $tyrA_{Eh}$ expression cassette resulting in the formation of pMON69924 (Figure 12).

A fourth plasmid was constructed by cloning the Arcelin 5-expression cassette for *slr1736* (SEQ ID NO: 29), downstream of the $HPPD_{At}$, and the *tyrA*$_{Eh}$ expression cassettes, resulting in the formation of pMON69943 (Figure 13).

Each of these binary constructs was transformed into soybean. R1 seed pools from plants harboring these constructs were analyzed for tocopherol content and composition. For constructs pMON36581 and pMON69933, the seed for analysis were chosen at random. Seed from plants transformed with pMON69924 and pMON69943 showed a segregating dark phenotype. This phenotype has been associated with the presence of increased levels of homogentisic acid as a result of the expression of trans genes *HPPD* and *tyrA*. Seed without dark coloration did have wild-type tocopherol levels and were not transgenic. For this reason colored seed were chosen for analysis of plants transformed with pMON69924 or pMON69943. For the impact of the *HPT* expression on total tocopherol accumulation in a single gene vector, or in a multi gene vector, seed from non-transformed soy, or seed transformed with pMON69924 served as controls, respectively. Figure 14 summarizes the tocopherol data obtained from these experiments. While expression of *ATPT2* or *slr1736* increased total tocopherol and tocotrienol levels in soy moderately, the impact of *HPT* expression in the context of a multi gene vector was much more pronounced. Figure 14 demonstrates a significantly increased level of tocopherol and tocotrienol accumulation for pMON69943 compared to pMON69924 lines. These data suggest that combination of an *HPT* with *tyrA*, and *HPPD* can substantially enhance tocopherol biosynthesis in soy.

Western analysis is carried out to detect the transgene expression in tissues harboring the gene of interest (GOI) expression cassette using the GOI protein specific antibody.

Northern analysis is done for detecting the mRNA level of the transgene using the GOI sequence specific radiolabelled probe.

EXAMPLE 7: Identification of additional homogentisate prenyl transferase sequences

In an analysis of the non-redundant amino acid database, Motifs II and IV (SEQ ID NOs: 40 and 42 identified in addition to *HPT* sequences, two genomic variants of *Arabidopsis thaliana* sequence related to *HPT*s (SEQ ID NOs: 61-62). These sequences are based on insillico prediction from genomic sequence by gene prediction algorithms. Further bioinformatic analysis showed that these sequences encoded an additional homogentisate prenyl transferase related to *HPT*. Both sequences (SEQ ID NOs: 61-62) were used to search the non-redundant amino acid database. The BLAST search results indicated that these sequences are related most to *HPT* sequences from cyanobacteria (SEQ ID NOs: 1-3) and *Arabidopsis* (SEQ ID NO: 7).

Alignment of gi15229898 (970 aa)(SEQ ID NO: 61) and gi10998133 (441 aa) (SEQ ID NO: 62) showed that:

a) C terminal half of gi15229898 (SEQ ID NO: 61) overlaps with gi10998133 (SEQ ID NO: 62);

b) the last 40-50 aa in the C terminal portions of these two proteins do not align; and

c) the N terminal of gi 15229898 does not align also with *HPT*s (SEQ ID NOs: 1-7, and 9-11). These findings indicate discrepancy in the coding sequence prediction reported in Genbank.

In order to verify the predicted sequence, the BAC sequence of the *Arabidopsis* genome corresponding to the region was downloaded from Genbank (gi|12408742|gb|AC016795.6|ATAC016795, 100835bp). Coding sequences were predicted from this BAC clone using the FGENESH (Solovyev V.V. (2001) Statistical approaches in Eukaryotic gene prediction: in Handbook of Statistical genetics (eds. Balding D. *et al.*), John Wiley & Sons, Ltd., p. 83-127) gene prediction program. FGENESH predicted 28 proteins from this BAC clone. To identify new homogentisate prenyl transferase proteins among 28 FGENESH predicted proteins, all 28 predicted proteins were blasted against the non-redundant amino acid database. FGENESH predicted protein No. 25 (402aa) (SEQ ID NO: 45) was most similar to gi10998133 (441 aa) (SEQ ID NO: 62), C terminal half of gi15229898 (970 aa) (SEQ ID NO: 61) and *HPT*s (SEQ ID NOs: 1-7, and 9-11.)

To provide functional and transcriptional evidence and to confirm the coding sequence for this gene, plant EST sequences database comprising proprietary and public sequence was searched. We found several ESTs (SEQ ID NOs: 63-72) which match the N terminal and C terminal portions of this gene. The new gene was named *HPT2* (SEQ ID NO: 59) from *Arabidopsis*. The *HPT2* (SEQ ID NO: 57) sequence is quite distinct from HPT1 (SEQ ID NO: 7).

HPT2 (SEQ ID NO: 57) from *Arabidopsis* is also known as Tocopherol Synthase (TS). Present data suggests that the overexpression of TS leads to a similar increase in the amount of overall tocopherol, over the wild type, as with HPT1 (SEQ ID NO: 33). However, the enzymes may have different biochemical characteristics because the overexpression of TS results in less production of the delta tocopherol than the overexpression of HPT1 (SEQ ID NO: 33).

The presence of chloroplast transit peptide in the HPT2 *Arabidopsis* sequence (SEQ ID NOs: 45 and 57) was verified using ChloroP program (Olof Emanuelsson1, Henrik Nielsen1, 2, and Gunnar von Heijne1ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science: 8: 978-984, 1999).

In addition to SEQ ID NOs: 1, 7, and 9-11(*HPT*), SEQ ID NOs: 57-58, and 90 (HPT2) were added to the alignment, see Figures 24-25 and the resulting motifs analyzed. Motif V (SEQ ID NO: 46), VII (SEQ ID NO: 48), and VIII (SEQ ID NO: 49) are specific to *HPT* and HPT2 sequences. A HMM search of the non-redundant amino acid database using these motifs identified only cyanobacteria (SEQ ID NOs: 1-3, and 43), photobacteria (SEQ ID NO: 44), and plant *HPT*s (SEQ ID NOs: 7, and 61-62). Motif VII (SEQ ID NO: 48) identified distantly related ubiA prenyl transerase from bacteria in addtion to homogentisate prenyl transferase. However, the sensitivity of Motif VII to homogentisate prenyl transferase was higher. Homogentisate prenyl transferases had lower e-values by several orders and hihger alignment score (higher than 30). HPT2 sequences are distinct from *HPT* and cyanobacterial *HPT*s as demonstrated by the sequence dendogram in Figure 26.

SEQ ID NOs: 43-44 were added to an alignment with SEQ ID NOs: 1-4, 6-7, 9-11, 57-58, and 91, see Figures 33-34, and the resulting motifs (SEQ ID NOs: 92-95, Motifs IX-XII) were analyzed. Specificity of these motifs to homogentisate prenyl transferases was confirmed by HMM search. A non redundant database containing more than 1.34M sequence was searched using HMM models built from the alignments shown in Figure 34 for Motifs

IX- XII. E value limits for the search was set at 1.0. All four motifs identified only homogentisate prenyl transferase from cyanobacteria, photobacteria and *Arabidopsis*. Upper E values limits for Motif IX, X, XI, and XII were 0.9, $11E10^{-11}$, 0.03, $8E10^{-8}$ respectively. The small size of motifs resulted in higher E values for Motif IX and XI.

EXAMPLE 8: Transformation and expression of a wild type *Arabidopsis* HPT2 gene in sense and antisense orientations with respect to seed-specific and constitutive promoters in *Arabidopsis thialiana*.

The HPT2 full-length cDNA (SEQ ID NO: 59) is excised from an EST clone, CPR230005 (pMON69960-Figure 15), with SalI and NotI enzymes, blunt-ended and cloned in between the napin promoter and napin 3' end at blunt-ended SalI site in sense and antisense orientations with respect to the napin promoter in pMON36525 (Figure 16) to generate recombinant binary vectors pMON69963 (Figure 17) and pMON69965 (Figure 18), respectively. The sequence of the HPT2 cDNA is confirmed by sequencing with napin 5'-sense (5'-GTGGCTCGGCTTCACTTTTTAC-3') (SEQ ID NO: 50) and napin 3'-antisense (5'-CCACACTCATATCACCGTGG-3') (SEQ ID NO: 51) primers using standard sequencing methodology. The HPT2 cDNA used to generate the pMON69963 and pMON69965 is also cloned in between the enhanced 35S promoter and E9-3' end at blunt-ended BglII and BamHI sites of pMON10098 (Figure 19) to generate the pMON69964 (Figure 20) and pMON69966 (Figure 21) in sense and antisense orientations with respect to the enhanced 35S promoter, respectively. Additional HPT2 internal primers synthesized to completely sequence the whole HPT2 cDNA are listed in the table below:

A list of primers used for confirming the HPT2 cDNA sequence.

Primer	Description	Sequence
BXK169	HPT2/CPR23005/sense	5'-CAGTGCTGGATAGAATTGCCCGGTTCC-3' (SEQ ID NO: 52)
BXK170	HPT2/CPR23005/sense	5'-GAGATCTATCAGTGCAGTCTGCTTGG-3' (SEQ ID NO: 53)
BXK171	HPT2/CPR23005/antisense	5'-GGGACAAGCATTTTTATTGCAAG-3' (SEQ ID NO: 54)
BXK172	HPT2/CPR23005/antisense	5'-GCCAAGATCACATGTGCAGGAATC-3' (SEQ ID NO: 55)
BXK173	HPT2/CPR23005/sense	5'-GTGGAGTGCACCTGTGGCGTTCATC-3' (SEQ ID NO: 56)

The plant binary vectors pMON69963, and pMON69965 are used in *Arabidopsis thaliana* plant transformation to direct the sense and antisense expression of the HPT2, in the

embryo. The binary vectors pMON69964, and pMON69966 are used in *Arabidopsis thaliana* plant transformation for sense and antisense expression of the HPT2 in whole plant. The binary vectors are transformed into ABI strain *Agrobacterium* cells by electroporation (Bio-Rad Electroprotocol Manual, Dower *et al., Nucleic Acids Res.*, 16:6127-6145 (1988)). Transgenic *Arabidopsis thaliana* plants are obtained by *Agrobacterium*-mediated transformation as described by Valverkens *et al., Proc. Nat. Acad. Sci., 85*:5536-5540 (1988), Bent *et al., Science,* 265:1856-1860 (1994), and Bechtold *et al., C.R. Acad. Sci., Life Sciences,* 316:1194-1199 (1993). Transgenic plants are selected by sprinkling the transformed T_1 seeds onto the selection plates containing MS basal salts (4.3 g/L), Gamborg'a B-5, 500X (2.0 g/L), sucrose (10 g/L), MES (0.5 g/L), phytagar (8 g/L), carbenicillin (250 mg/L), cefotaxime (100 mg/L), plant preservation medium (2 ml/L), and kanamycin (60 mg/L) and then vernalizing them at 4°C in the absence of light for 2-4 days. The seeds are transferred to 23°C, and 16/8 hours light/dark cycle for 5-10 days until seedlings emerge. Once one set of true leaves are formed on the kanamycin resistant seedlings, they are transferred to soil and grown to maturity. The transgenic lines generated through kanamycin selection are grown under two different light conditions. One set of the transgenic lines are grown under 16 hrs light and 8 hrs dark and another set of the transgenic lines are grown under 24 hrs light to study the effect of light on seed tocopherol levels. The T_2 seed harvested from the transformants is analyzed for tocopherol content. The results from the seed total tocopherol analysis from lines grown under both normal and high light conditions are presented in Figures 22 and 23. Seed-specific overexpression of HPT2 under normal and high light conditions produced a significant 1.6- and 1.5-fold increase in total tocopherol levels (alpha=0.05; *Tukey-Kramer HSD)* (SAS institute, 2002, JPM version 5.0).

Expression of HPT2 using the constitutive promoter, e35S, produced about 20% increase in seed total tocopherol levels as compared to control under both light conditions. Maximum tocopherol level reduction in lines harboring the enhanced 35S::HPT2 antisense construct was 20%. Overall, the significant increase in seed total tocopherol level in the *Arabidopsis thaliana* lines harboring the HPT2 driven by the napin promoter suggests that HPT2 plays a key role in tocopherol biosynthesis.

Western analysis is carried out to detect the transgene expression in tissues harboring the gene of interest (GOI) expression cassette using the GOI protein specific antibody.

Northern analysis is done for detecting the mRNA level of the transgene using the GOI sequence specific radiolabelled probe.

EXAMPLE 9: Preparation of plant binary vector for expression of HPT2 from *Arabidopsis* in combination with tocopherol pathway genes

To investigate the combinatorial effect of HPT2 with other key enzymes in the pathway, a plant binary vector containing seed-specifically expressed hydroxyphenylpyruvate dioxygenase (*HPPD*), bifunctional prephenate dehydrogenase *tyrA*, and HPT2 (pMON81028 – Figure 27) is prepared. The pMON81028 is made by exercising the pNapin::HPT2::Napin 3' expression cassette from pMON81023 (Figure 28) with Bsp120I and NotI enzymes and ligating it to pMON36596 (Figure 29) at NotI site. The pMON36596 contains the pNapin::CTP2::HPPD::Napin 3' and pNapin::CTP1::TyrA::Napin 3' expression cassettes. The pMON81028 is transformed into *Arabidopsis thaliana* plant using the method described in Example 8.

EXAMPLE 10: Preparation of construct for bacterial expression of HPT2 from *Arabidopsis*

The EST clone CPR23005 containing the HPT2 full length cDNA is used as a template for PCR to amplify the HPT2 cDNA fragment codes for the mature form of the HPT2 protein. Two sets of PCR products are generated to clone at the pET30a(+) vector (Novagen, Inc.) (Figure 30) to produce HPT2 protein with and without his tag. The primer set BXK174 (5'- CACATATGGCATGTTCTCAGGTTGGTGCTGC-3') (SEQ ID NO: 84) and BXK176 (5'-GCGTCGACCTAGAGGAAGGGGAATAACAG-3') (SEQ ID NO: 85) is used for cloning HPT2 at the NdeI and SalI sites of pET30a(+), behind the T7 promoter to generate mature HPT2 protein without the his tag. The resulting recobmbinant vector is named pMON69993 (Figure 31). The primer set BXK175 (5'-CAACCATGGCATGTTCTCAGGTTGGTGCTGC-3') (SEQ ID NO: 86) and BXK176 (5'-GCGTCGACCTAGAGGAAGGGGAATAACAG-3') (SEQ ID NO: 87) is used to generate HPT2 PCR product to clone at the NcoI and SalI sites of pET30a(+) to produce mature HPT2 with his tag. The recombinant vector is named as pMON69992 (Figure 32). The pMON69993 and pMON69992 is used for producing bacterial expressed HPT2 to carry out enzyme assays to confirm its homogentisate prenyl transferase activity and specificity towards geranylgeranyl pyrophosphosphate, phytyl pyrophophaste and solanyl pyrophosphate substrates.

SEQUENCE LISTING

<110> Monsanto Technology LLC
Valentin, Henry E
Venkatesh, Tyamagondlu V
Karunanandaa, Balasulojini

<120> Homogentisate Prenyl Transferase ("HPT") Nucleic Acids and Polypeptides, and Uses Thereof

<130> REN-02-052 MON-52933

<150> 60/365,202
<151> 2002-03-19

<160> 95

<170> PatentIn version 3.1

<210> 1
<211> 322
<212> PRT
<213> Nostoc punctiforme

<400> 1

Met Ser Gln Ser Ser Gln Asn Ser Pro Leu Pro Arg Lys Pro Val Gln
1 5 10 15

Ser Tyr Phe His Trp Leu Tyr Ala Phe Trp Lys Phe Ser Arg Pro His
20 25 30

Thr Ile Ile Gly Thr Ser Leu Ser Val Leu Ser Leu Tyr Leu Ile Ala
35 40 45

Ile Ala Ile Ser Asn Asn Thr Ala Ser Leu Phe Thr Thr Pro Gly Ser
50 55 60

Leu Ser Pro Leu Phe Gly Ala Trp Ile Ala Cys Leu Cys Gly Asn Val
65 70 75 80

Tyr Ile Val Gly Leu Asn Gln Leu Glu Asp Val Asp Ile Asp Lys Ile
85 90 95

Asn Lys Pro His Leu Pro Leu Ala Ser Gly Glu Phe Ser Gln Gln Thr
100 105 110

Gly Gln Leu Ile Val Ala Ser Thr Gly Ile Leu Ala Leu Val Met Ala
115 120 125

Trp Leu Thr Gly Pro Phe Leu Phe Gly Met Val Thr Ile Ser Leu Ala
130 135 140

Ile Gly Thr Ala Tyr Ser Leu Pro Pro Ile Arg Leu Lys Gln Phe Pro
145 150 155 160

Phe Trp Ala Ala Leu Cys Ile Phe Ser Val Arg Gly Thr Ile Val Asn
165 170 175

Leu Gly Leu Tyr Leu His Tyr Ser Trp Ala Leu Lys Gln Ser Gln Thr
180 185 190

Ile Pro Pro Val Val Trp Val Leu Thr Leu Phe Ile Leu Val Phe Thr
195 200 205

Phe Ala Ile Ala Ile Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Arg
210 215 220

Leu Tyr Asn Ile Thr Thr Phe Thr Ile Lys Leu Gly Ser Gln Ala Val
225 230 235 240

Phe Asn Leu Ala Leu Trp Val Ile Thr Val Cys Tyr Leu Gly Ile Ile
245 250 255

Leu Val Gly Val Leu Arg Ile Ala Ser Val Asn Pro Ile Phe Leu Ile
260 265 270

Thr Ala His Leu Ala Leu Leu Val Trp Met Trp Trp Arg Ser Leu Ala
275 280 285

Val Asp Leu Gln Asp Lys Ser Ala Ile Ala Gln Phe Tyr Gln Phe Ile
290 295 300

Trp Lys Leu Phe Phe Ile Glu Tyr Leu Ile Phe Pro Ile Ala Cys Phe
305 310 315 320

Leu Ala

<210> 2
<211> 318
<212> PRT
<213> Anabaena sp.

<400> 2

Met Asn Gln Ser Ser Gln Asp Arg Pro Leu Arg Pro Lys Pro Leu Gln
1 5 10 15

Ser Ser Phe Gln Trp Leu Tyr Ala Phe Trp Lys Phe Ser Arg Pro His

20 25 30

Thr Ile Ile Gly Thr Ser Leu Ser Val Leu Gly Leu Tyr Leu Ile Ser
35 40 45

Ile Ala Val Ser Ser Thr Gly Phe Ala Leu Thr Gln Ile Asn Ser Val
50 55 60

Leu Gly Ala Trp Leu Ala Cys Leu Cys Gly Asn Val Tyr Ile Val Gly
65 70 75 80

Leu Asn Gln Leu Glu Asp Ile Glu Ile Asp Lys Val Asn Lys Pro His
85 90 95

Leu Pro Leu Ala Ser Gly Glu Phe Ser Arg Lys Gln Gly Arg Ile Ile
100 105 110

Val Ile Leu Thr Gly Ile Thr Ala Ile Val Leu Ala Trp Leu Asn Gly
115 120 125

Pro Tyr Leu Phe Gly Met Val Ala Val Ser Leu Ala Ile Gly Thr Ala
130 135 140

Tyr Ser Leu Pro Pro Ile Arg Leu Lys Gln Phe Pro Phe Trp Ala Ala
145 150 155 160

Leu Cys Ile Phe Ser Val Arg Gly Thr Ile Val Asn Leu Gly Leu Tyr
165 170 175

Leu His Phe Ser Trp Leu Leu Gln Asn Lys Gln Ser Ile Pro Leu Pro
180 185 190

Val Trp Ile Leu Thr Val Phe Ile Leu Ile Phe Thr Phe Ala Ile Ala
195 200 205

Ile Phe Lys Asp Ile Pro Asp Met Glu Gly Asp Arg Leu Tyr Asn Ile
210 215 220

Thr Thr Leu Thr Ile Gln Leu Gly Pro Gln Ala Val Phe Asn Leu Ala
225 230 235 240

Met Trp Val Leu Thr Val Cys Tyr Leu Gly Met Val Ile Ile Gly Val
245 250 255

Leu Arg Leu Gly Thr Ile Asn Ser Val Phe Leu Val Val Thr His Leu
260 265 270

Val Ile Leu Cys Trp Met Trp Met Gln Ser Leu Ala Val Asp Ile His
275 280 285

Asp Lys Thr Ala Ile Ala Gln Phe Tyr Gln Phe Ile Trp Lys Leu Phe
290 295 300

Phe Leu Glu Tyr Leu Met Phe Pro Ile Ala Cys Leu Leu Ala
305 310 315

<210> 3
<211> 308
<212> PRT
<213> Synechocystis sp.

<400> 3

Met Ala Thr Ile Gln Ala Phe Trp Arg Phe Ser Arg Pro His Thr Ile
1 5 10 15

Ile Gly Thr Thr Leu Ser Val Trp Ala Val Tyr Leu Leu Thr Ile Leu
20 25 30

Gly Asp Gly Asn Ser Val Asn Ser Pro Ala Ser Leu Asp Leu Val Phe
35 40 45

Gly Ala Trp Leu Ala Cys Leu Leu Gly Asn Val Tyr Ile Val Gly Leu
50 55 60

Asn Gln Leu Trp Asp Val Asp Ile Asp Arg Ile Asn Lys Pro Asn Leu
65 70 75 80

Pro Leu Ala Asn Gly Asp Phe Ser Ile Ala Gln Gly Arg Trp Ile Val
85 90 95

Gly Leu Cys Gly Val Ala Ser Leu Ala Ile Ala Trp Gly Leu Gly Leu
100 105 110

Trp Leu Gly Leu Thr Val Gly Ile Ser Leu Ile Ile Gly Thr Ala Tyr
115 120 125

Ser Val Pro Pro Val Arg Leu Lys Arg Phe Ser Leu Leu Ala Ala Leu
130 135 140

Cys Ile Leu Thr Val Arg Gly Ile Val Val Asn Leu Gly Leu Phe Leu
145 150 155 160

Phe Phe Arg Ile Gly Leu Gly Tyr Pro Pro Thr Leu Ile Thr Pro Ile
165 170 175

Trp Val Leu Thr Leu Phe Ile Leu Val Phe Thr Val Ala Ile Ala Ile
180 185 190

Phe Lys Asp Val Pro Asp Met Glu Gly Asp Arg Gln Phe Lys Ile Gln
195 200 205

Thr Leu Thr Leu Gln Ile Gly Lys Gln Asn Val Phe Arg Gly Thr Leu
210 215 220

Ile Leu Leu Thr Gly Cys Tyr Leu Ala Met Ala Ile Trp Gly Leu Trp
225 230 235 240

Ala Ala Met Pro Leu Asn Thr Ala Phe Leu Ile Val Ser His Leu Cys
245 250 255

Leu Leu Ala Leu Leu Trp Trp Arg Ser Arg Asp Val His Leu Glu Ser
260 265 270

Lys Thr Glu Ile Ala Ser Phe Tyr Gln Phe Ile Trp Lys Leu Phe Phe
275 280 285

Leu Glu Tyr Leu Leu Tyr Pro Leu Ala Leu Trp Leu Pro Asn Phe Ser
290 295 300

Asn Thr Ile Phe
305

<210> 4
<211> 399
<212> PRT
<213> Zea mays

<400> 4

Met Asp Ala Leu Arg Leu Arg Pro Ser Leu Leu Pro Val Arg Pro Gly
1 5 10 15

Ala Ala Arg Pro Arg Asp His Phe Leu Pro Pro Cys Cys Ser Ile Gln
20 25 30

Arg Asn Gly Glu Gly Arg Ile Cys Phe Ser Ser Gln Arg Thr Gln Gly
35 40 45

Pro Thr Leu His His His Gln Lys Phe Phe Glu Trp Lys Ser Ser Tyr
50 55 60

Cys Arg Ile Ser His Arg Ser Leu Asn Thr Ser Val Asn Ala Ser Gly
65 70 75 80

Gln Gln Leu Gln Ser Glu Pro Glu Thr His Asp Ser Thr Thr Ile Trp
85 90 95

Arg Ala Ile Ser Ser Ser Leu Asp Ala Phe Tyr Arg Phe Ser Arg Pro
100 105 110

His Thr Val Ile Gly Thr Ala Leu Ser Ile Val Ser Val Ser Leu Leu
115 120 125

Ala Val Gln Ser Leu Ser Asp Ile Ser Pro Leu Phe Leu Thr Gly Leu
130 135 140

Leu Glu Ala Val Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val Gly
145 150 155 160

Leu Asn Gln Leu Phe Asp Ile Glu Ile Asp Lys Val Asn Lys Pro Thr
165 170 175

Leu Pro Leu Ala Ser Gly Glu Tyr Thr Leu Ala Thr Gly Val Ala Ile
180 185 190

Val Ser Val Phe Ala Ala Met Ser Phe Gly Leu Gly Trp Ala Val Gly
195 200 205

Ser Gln Pro Leu Phe Trp Ala Leu Phe Ile Ser Phe Val Leu Gly Thr
210 215 220

Ala Tyr Ser Ile Asn Leu Pro Tyr Leu Arg Trp Lys Arg Phe Ala Val
225 230 235 240

Val Ala Ala Leu Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu
245 250 255

Ala Phe Phe Leu His Ile Gln Thr Phe Val Phe Arg Arg Pro Ala Val
260 265 270

Phe Ser Arg Pro Leu Leu Phe Ala Thr Gly Phe Met Thr Phe Phe Ser
275 280 285

Val Val Ile Ala Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Arg
290 295 300

Ile Phe Gly Ile Arg Ser Phe Ser Val Arg Leu Gly Gln Lys Lys Val
305 310 315 320

Phe Trp Ile Cys Val Gly Leu Leu Glu Met Ala Tyr Ser Val Ala Ile
325 330 335

Leu Met Gly Ala Thr Ser Ser Cys Leu Trp Ser Lys Thr Ala Thr Ile
340 345 350

Ala Gly His Ser Ile Leu Ala Ala Ile Leu Trp Ser Cys Ala Arg Ser
355 360 365

Val Asp Leu Thr Ser Lys Ala Ala Ile Thr Ser Phe Tyr Met Phe Ile
370 375 380

Trp Lys Leu Phe Tyr Ala Glu Tyr Leu Leu Ile Pro Leu Val Arg
385 390 395

<210> 5
<211> 411
<212> PRT
<213> Glycine max (ppt2)

<400> 5

Met Asp Ser Leu Leu Leu Arg Ser Phe Pro Asn Ile Asn Asn Ala Ser
1 5 10 15

Ser Leu Thr Thr Thr Gly Ala Asn Phe Ser Arg Thr Lys Ser Phe Ala
20 25 30

Asn Ile Tyr His Ala Ser Ser Tyr Val Pro Asn Ala Ser Trp His Asn
35 40 45

Arg Lys Ile Gln Lys Glu Tyr Asn Phe Leu Arg Phe Arg Trp Pro Ser
50 55 60

Leu Asn His His Tyr Lys Gly Ile Glu Gly Ala Cys Thr Cys Lys Lys
65 70 75 80

Cys Asn Ile Lys Phe Val Val Lys Ala Thr Ser Glu Lys Ser Leu Glu
85 90 95

Ser Glu Pro Gln Ala Phe Asp Pro Lys Ser Ile Leu Asp Ser Val Lys
100 105 110

Asn Ser Leu Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Val Ile

115 120 125

Gly Thr Ala Leu Ser Ile Ile Ser Val Ser Leu Leu Ala Val Glu Lys
130 135 140

Ile Ser Asp Ile Ser Pro Leu Phe Phe Thr Gly Val Leu Glu Ala Val
145 150 155 160

Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu
165 170 175

Ser Asp Val Glu Ile Asp Lys Ile Asn Lys Pro Tyr Leu Pro Leu Ala
180 185 190

Ser Gly Glu Tyr Ser Phe Glu Thr Gly Val Thr Ile Val Ala Ser Phe
195 200 205

Ser Ile Leu Ser Phe Trp Leu Gly Trp Val Val Gly Ser Trp Pro Leu
210 215 220

Phe Trp Ala Leu Phe Val Ser Phe Val Leu Gly Thr Ala Tyr Ser Ile
225 230 235 240

Asn Val Pro Leu Leu Arg Trp Lys Arg Phe Ala Val Leu Ala Ala Met
245 250 255

Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu Ala Phe Phe Leu
260 265 270

His Met Gln Thr His Val Tyr Lys Arg Pro Pro Val Phe Ser Arg Pro
275 280 285

Leu Ile Phe Ala Thr Ala Phe Met Ser Phe Phe Ser Val Val Ile Ala
290 295 300

Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Val Phe Gly Ile
305 310 315 320

Gln Ser Phe Ser Val Arg Leu Gly Gln Lys Pro Val Phe Trp Thr Cys
325 330 335

Val Thr Leu Leu Glu Ile Ala Tyr Gly Val Ala Leu Leu Val Gly Ala
340 345 350

Ala Ser Pro Cys Leu Trp Ser Lys Ile Phe Thr Gly Leu Gly His Ala
355 360 365

```
Val Leu Ala Ser Ile Leu Trp Phe His Ala Lys Ser Val Asp Leu Lys
    370                 375                 380

Ser Lys Ala Ser Ile Thr Ser Phe Tyr Met Phe Ile Trp Lys Leu Phe
385                 390                 395                 400

Tyr Ala Glu Tyr Leu Leu Ile Pro Phe Val Arg
                405                 410
```

<210> 6
<211> 395
<212> PRT
<213> Glycine max (ppt1)

<400> 6

```
Met Asp Ser Met Leu Leu Arg Ser Phe Pro Asn Ile Asn Asn Ala Ser
1               5                   10                  15

Ser Leu Ala Thr Thr Gly Ser Tyr Leu Pro Asn Ala Ser Trp His Asn
            20                  25                  30

Arg Lys Ile Gln Lys Glu Tyr Asn Phe Leu Arg Phe Arg Trp Pro Ser
        35                  40                  45

Leu Asn His His Tyr Lys Ser Ile Glu Gly Gly Cys Thr Cys Lys Lys
    50                  55                  60

Cys Asn Ile Lys Phe Val Val Lys Ala Thr Ser Glu Lys Ser Phe Glu
65                  70                  75                  80

Ser Glu Pro Gln Ala Phe Asp Pro Lys Ser Ile Leu Asp Ser Val Lys
                85                  90                  95

Asn Ser Leu Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Val Ile
            100                 105                 110

Gly Thr Ala Leu Ser Ile Ile Ser Val Ser Leu Leu Ala Val Glu Lys
        115                 120                 125

Ile Ser Asp Ile Ser Pro Leu Phe Phe Thr Gly Val Leu Glu Ala Val
    130                 135                 140

Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu
145                 150                 155                 160
```

```
Ser Asp Val Glu Ile Asp Lys Ile Asn Lys Pro Tyr Leu Pro Leu Ala
                165                 170                 175

Ser Gly Glu Tyr Ser Phe Glu Thr Gly Val Thr Ile Val Ala Ser Phe
            180                 185                 190

Ser Ile Leu Ser Phe Trp Leu Gly Trp Val Val Gly Ser Trp Pro Leu
        195                 200                 205

Phe Trp Ala Leu Phe Val Ser Phe Val Leu Gly Thr Ala Tyr Ser Ile
    210                 215                 220

Asn Val Pro Leu Leu Arg Trp Lys Arg Phe Ala Val Leu Ala Ala Met
225                 230                 235                 240

Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu Ala Phe Phe Leu
                245                 250                 255

His Ile Gln Thr His Val Tyr Lys Arg Pro Pro Val Phe Ser Arg Ser
            260                 265                 270

Leu Ile Phe Ala Thr Ala Phe Met Ser Phe Phe Ser Val Val Ile Ala
        275                 280                 285

Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Val Phe Gly Ile
    290                 295                 300

Gln Ser Phe Ser Val Arg Leu Gly Gln Lys Pro Val Phe Trp Thr Cys
305                 310                 315                 320

Val Ile Leu Leu Glu Ile Ala Tyr Gly Val Ala Leu Leu Val Gly Ala
                325                 330                 335

Ala Ser Pro Cys Leu Trp Ser Lys Ile Val Thr Gly Leu Gly His Ala
            340                 345                 350

Val Leu Ala Ser Ile Leu Trp Phe His Ala Lys Ser Val Asp Leu Lys
        355                 360                 365

Ser Lys Ala Ser Ile Thr Ser Phe Tyr Met Phe Ile Trp Lys Leu Phe
    370                 375                 380

Tyr Ala Glu Tyr Leu Leu Ile Pro Phe Val Arg
385                 390                 395
```

<210> 7
<211> 393
<212> PRT
<213> Arabidopsis thaliana

<400> 7

Met Glu Ser Leu Leu Ser Ser Ser Ser Leu Val Ser Ala Ala Gly Gly
1 5 10 15

Phe Cys Trp Lys Lys Gln Asn Leu Lys Leu His Ser Leu Ser Glu Ile
20 25 30

Arg Val Leu Arg Cys Asp Ser Ser Lys Val Val Ala Lys Pro Lys Phe
35 40 45

Arg Asn Asn Leu Val Arg Pro Asp Gly Gln Gly Ser Ser Leu Leu Leu
50 55 60

Tyr Pro Lys His Lys Ser Arg Phe Arg Val Asn Ala Thr Ala Gly Gln
65 70 75 80

Pro Glu Ala Phe Asp Ser Asn Ser Lys Gln Lys Ser Phe Arg Asp Ser
85 90 95

Leu Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Val Ile Gly Thr
100 105 110

Val Leu Ser Ile Leu Ser Val Ser Phe Leu Ala Val Glu Lys Val Ser
115 120 125

Asp Ile Ser Pro Leu Leu Phe Thr Gly Ile Leu Glu Ala Val Val Ala
130 135 140

Ala Leu Met Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu Ser Asp
145 150 155 160

Val Glu Ile Asp Lys Val Asn Lys Pro Tyr Leu Pro Leu Ala Ser Gly
165 170 175

Glu Tyr Ser Val Asn Thr Gly Ile Ala Ile Val Ala Ser Phe Ser Ile
180 185 190

Met Ser Phe Trp Leu Gly Trp Ile Val Gly Ser Trp Pro Leu Phe Trp
195 200 205

Ala Leu Phe Val Ser Phe Met Leu Gly Thr Ala Tyr Ser Ile Asn Leu
210 215 220

```
Pro Leu Leu Arg Trp Lys Arg Phe Ala Leu Val Ala Ala Met Cys Ile
225                 230                 235                 240

Leu Ala Val Arg Ala Ile Ile Val Gln Ile Ala Phe Tyr Leu His Ile
                245                 250                 255

Gln Thr His Val Phe Gly Arg Pro Ile Leu Phe Thr Arg Pro Leu Ile
            260                 265                 270

Phe Ala Thr Ala Phe Met Ser Phe Phe Ser Val Val Ile Ala Leu Phe
        275                 280                 285

Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Ile Phe Gly Ile Arg Ser
    290                 295                 300

Phe Ser Val Thr Leu Gly Gln Lys Arg Val Phe Trp Thr Cys Val Thr
305                 310                 315                 320

Leu Leu Gln Met Ala Tyr Ala Val Ala Ile Leu Val Gly Ala Thr Ser
                325                 330                 335

Pro Phe Ile Trp Ser Lys Val Ile Ser Val Val Gly His Val Ile Leu
            340                 345                 350

Ala Thr Thr Leu Trp Ala Arg Ala Lys Ser Val Asp Leu Ser Ser Lys
        355                 360                 365

Thr Glu Ile Thr Ser Cys Tyr Met Phe Ile Trp Lys Leu Phe Tyr Ala
    370                 375                 380

Glu Tyr Leu Leu Leu Pro Phe Leu Lys
385                 390
```

<210> 8
<211> 361
<212> PRT
<213> Cuphea pulcherrima

<400> 8

```
Met Arg Met Glu Ser Leu Leu Leu Asn Ser Phe Ser Pro Ser Pro Ala
1               5                   10                  15

Gly Gly Lys Ile Cys Arg Ala Asp Thr Tyr Lys Lys Ala Tyr Phe Ala
            20                  25                  30
```

Thr Ala Arg Cys Asn Thr Leu Asn Ser Leu Asn Lys Asn Thr Gly Glu
35 40 45

Tyr His Leu Ser Arg Thr Arg Gln Arg Phe Thr Phe His Gln Asn Gly
50 55 60

His Arg Thr Tyr Leu Val Lys Ala Val Ser Gly Gln Ser Leu Glu Ser
65 70 75 80

Glu Pro Glu Ser Tyr Pro Asn Asn Arg Trp Asp Tyr Val Lys Ser Ala
85 90 95

Ala Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Ile Ile Gly Thr
100 105 110

Ala Leu Ser Ile Val Ser Val Ser Leu Leu Ala Val Glu Lys Leu Pro
115 120 125

Glu Leu Asn Ser Met Phe Phe Thr Gly Leu Leu Glu Val Ile Leu Ala
130 135 140

Ala Leu Phe Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu Ser Asp
145 150 155 160

Ile Asp Ile Asp Lys Val Asn Lys Pro Tyr Leu Pro Leu Ala Ser Gly
165 170 175

Glu Phe Ser Val Gly Thr Gly Val Thr Ile Val Thr Ser Phe Leu Ile
180 185 190

Met Ser Phe Trp Leu Gly Trp Val Val Gly Ser Trp Pro Leu Phe Trp
195 200 205

Ala Leu Phe Ile Ser Phe Val Leu Gly Thr Ala Tyr Ser Ile Asp Met
210 215 220

Pro Met Leu Arg Trp Lys Arg Ser Ala Val Val Ala Ala Leu Cys Ile
225 230 235 240

Leu Ala Val Arg Ala Val Ile Val Gln Ile Ala Phe Phe Leu His Met
245 250 255

Gln Met His Val Tyr Gly Arg Ala Ala Ala Leu Ser Arg Pro Val Ile
260 265 270

Phe Ala Thr Gly Phe Met Ser Phe Phe Ser Ile Val Ile Ala Leu Phe

275 280 285

Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Ile Phe Gly Ile Arg Ser
290 295 300

Phe Thr Val Arg Leu Gly Gln Glu Arg Val Phe Trp Ile Cys Ile Ser
305 310 315 320

Leu Leu Glu Met Ala Tyr Ala Val Ala Leu Trp Val Leu Arg Ala Arg
325 330 335

Gly Arg Lys Lys His Ala Asp Gly Val Ser Ala Ser Glu Phe Phe Leu
340 345 350

Ser Ile Ser Gly Gly Arg Lys Asn Leu
355 360

<210> 9
<211> 395
<212> PRT
<213> Allium porrum

<400> 9

Met Leu Ser Met Asp Ser Leu Leu Thr Lys Pro Val Val Ile Pro Leu
1 5 10 15

Pro Ser Pro Val Cys Ser Leu Pro Ile Leu Arg Gly Ser Ser Ala Pro
20 25 30

Gly Gln Tyr Ser Cys Arg Asn Tyr Asn Pro Ile Arg Ile Gln Arg Cys
35 40 45

Leu Val Asn Tyr Glu His Val Lys Pro Arg Phe Thr Thr Cys Ser Arg
50 55 60

Ser Gln Lys Leu Gly His Val Lys Ala Thr Ser Glu His Ser Leu Glu
65 70 75 80

Ser Gly Ser Glu Gly Tyr Thr Pro Arg Ser Ile Trp Glu Ala Val Leu
85 90 95

Ala Ser Leu Asn Val Leu Tyr Lys Phe Ser Arg Pro His Thr Ile Ile
100 105 110

Gly Thr Ala Met Gly Ile Met Ser Val Ser Leu Leu Val Val Glu Ser
115 120 125

```
Leu Ser Asp Ile Ser Pro Leu Phe Phe Val Gly Leu Leu Glu Ala Val
    130                 135                 140

Val Ala Ala Leu Phe Met Asn Val Tyr Ile Val Gly Leu Asn Gln Leu
145                 150                 155                 160

Phe Asp Ile Glu Ile Asp Lys Val Asn Lys Pro Asp Leu Pro Leu Ala
                165                 170                 175

Ser Gly Glu Tyr Ser Pro Arg Ala Gly Thr Ala Ile Val Ile Ala Ser
            180                 185                 190

Ala Ile Met Ser Phe Gly Ile Gly Trp Leu Val Gly Ser Trp Pro Leu
        195                 200                 205

Phe Trp Ala Leu Phe Ile Ser Phe Val Leu Gly Thr Ala Tyr Ser Ile
    210                 215                 220

Asn Leu Pro Phe Leu Arg Trp Lys Arg Ser Ala Val Val Ala Ala Ile
225                 230                 235                 240

Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu Ala Phe Phe Leu
                245                 250                 255

His Ile Gln Ser Phe Val Phe Lys Arg Pro Ala Ser Phe Thr Arg Pro
            260                 265                 270

Leu Ile Phe Ala Thr Ala Phe Met Ser Phe Phe Ser Val Val Ile Ala
        275                 280                 285

Leu Phe Lys Asp Ile Pro Asp Ile Asp Gly Asp Lys Ile Phe Gly Ile
    290                 295                 300

His Ser Phe Ser Val Arg Leu Gly Gln Glu Arg Val Phe Trp Ile Cys
305                 310                 315                 320

Ile Tyr Leu Leu Glu Met Ala Tyr Thr Val Val Met Val Val Gly Ala
                325                 330                 335

Thr Ser Ser Cys Leu Trp Ser Lys Cys Leu Thr Val Ile Gly His Ala
            340                 345                 350

Ile Leu Gly Ser Leu Leu Trp Asn Arg Ala Arg Ser His Gly Pro Met
        355                 360                 365
```

Thr Lys Thr Thr Ile Thr Ser Phe Tyr Met Phe Val Trp Lys Leu Phe
370 375 380

Tyr Ala Glu Tyr Leu Leu Ile Pro Phe Val Arg
385 390 395

<210> 10
<211> 400
<212> PRT
<213> Triticum sp.

<400> 10

Met Asp Ser Leu Arg Leu Arg Pro Ser Ser Leu Arg Ser Ala Pro Gly
1 5 10 15

Ala Ala Ala Ala Arg Arg Arg Asp His Ile Leu Pro Ser Phe Cys Ser
20 25 30

Ile Gln Arg Asn Gly Lys Gly Arg Val Thr Leu Ser Ile Gln Ala Ser
35 40 45

Lys Gly Pro Thr Ile Asn His Cys Lys Lys Phe Leu Asp Trp Lys Tyr
50 55 60

Ser Asn His Arg Ile Ser His Gln Ser Ile Asn Thr Ser Ala Lys Ala
65 70 75 80

Gly Gln Ser Leu Gln Pro Glu Thr Glu Ala His Asp Pro Ala Ser Phe
85 90 95

Trp Lys Pro Ile Ser Ser Ser Leu Asp Ala Phe Tyr Arg Phe Ser Arg
100 105 110

Pro His Thr Ile Ile Gly Thr Ala Leu Ser Ile Val Ser Val Ser Leu
115 120 125

Leu Ala Val Glu Ser Leu Ser Asp Ile Ser Pro Leu Phe Leu Thr Gly
130 135 140

Leu Leu Glu Ala Val Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val
145 150 155 160

Gly Leu Asn Gln Leu Phe Asp Ile Glu Ile Asp Lys Val Asn Lys Pro
165 170 175

Thr Leu Pro Leu Ala Ser Gly Glu Tyr Ser Pro Ala Thr Gly Val Ala
180 185 190

Ile Val Ser Val Phe Ala Ala Met Ser Phe Gly Leu Gly Trp Val Val
195 200 205

Gly Ser Pro Pro Leu Phe Trp Ala Leu Phe Ile Ser Phe Val Leu Gly
210 215 220

Thr Ala Tyr Ser Val Asn Leu Pro Tyr Phe Arg Trp Lys Arg Ser Ala
225 230 235 240

Val Val Ala Ala Leu Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln
245 250 255

Leu Ala Phe Phe Leu His Ile Gln Thr Phe Val Phe Arg Arg Pro Ala
260 265 270

Val Phe Ser Lys Pro Leu Ile Phe Ala Thr Ala Phe Met Thr Phe Phe
275 280 285

Ser Val Val Ile Ala Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp
290 295 300

Arg Ile Phe Gly Ile Gln Ser Phe Ser Val Arg Leu Gly Gln Ser Lys
305 310 315 320

Val Phe Trp Thr Cys Val Gly Leu Leu Glu Val Ala Tyr Gly Val Ala
325 330 335

Ile Leu Met Gly Val Thr Ser Ser Ser Leu Trp Ser Lys Ser Leu Thr
340 345 350

Val Val Gly His Ala Ile Leu Ala Ser Ile Leu Trp Ser Ser Ala Arg
355 360 365

Ser Ile Asp Leu Thr Ser Lys Ala Ala Ile Thr Ser Phe Tyr Met Leu
370 375 380

Ile Trp Arg Leu Phe Tyr Ala Glu Tyr Leu Leu Ile Pro Leu Val Arg
385 390 395 400

<210> 11
<211> 393
<212> PRT
<213> Cuphea pulcherrima

<400> 11

```
Met Arg Met Glu Ser Leu Leu Leu Asn Ser Phe Ser Pro Ser Pro Ala
1               5                  10                  15

Gly Gly Lys Ile Cys Arg Ala Asp Thr Tyr Lys Lys Ala Tyr Phe Ala
            20                  25                  30

Thr Ala Arg Cys Asn Thr Leu Asn Ser Leu Asn Lys Asn Thr Gly Glu
        35                  40                  45

Tyr His Leu Ser Arg Thr Arg Gln Arg Phe Thr Phe His Gln Asn Gly
    50                  55                  60

His Arg Thr Tyr Leu Val Lys Ala Val Ser Gly Gln Ser Leu Glu Ser
65                  70                  75                  80

Glu Pro Glu Ser Tyr Pro Asn Asn Arg Trp Asp Tyr Val Lys Ser Ala
                85                  90                  95

Ala Asp Ala Phe Tyr Arg Phe Ser Arg Pro His Thr Ile Ile Gly Thr
            100                 105                 110

Ala Leu Ser Ile Val Ser Val Ser Leu Leu Ala Val Glu Lys Leu Pro
        115                 120                 125

Glu Leu Asn Ser Met Phe Phe Thr Gly Leu Leu Glu Val Ile Leu Ala
    130                 135                 140

Ala Leu Phe Met Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu Ser Asp
145                 150                 155                 160

Ile Asp Ile Asp Lys Val Asn Lys Pro Tyr Leu Pro Leu Ala Ser Gly
                165                 170                 175

Glu Phe Ser Val Gly Thr Gly Val Thr Ile Val Thr Ser Phe Leu Ile
            180                 185                 190

Met Ser Phe Trp Leu Gly Trp Val Val Gly Ser Trp Pro Leu Phe Trp
        195                 200                 205

Ala Leu Phe Ile Ser Phe Val Leu Gly Thr Ala Tyr Ser Ile Asp Met
    210                 215                 220

Pro Met Leu Arg Trp Lys Arg Ser Ala Val Val Ala Ala Leu Cys Ile
225                 230                 235                 240

Leu Ala Val Arg Ala Val Ile Val Gln Ile Ala Phe Phe Leu His Met
```

245 250 255

Gln Met His Val Tyr Gly Arg Ala Ala Ala Leu Ser Arg Pro Val Ile
260 265 270

Phe Ala Thr Gly Phe Met Ser Phe Phe Ser Ile Val Ile Ala Leu Phe
275 280 285

Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys Ile Phe Gly Ile Arg Ser
290 295 300

Phe Thr Val Arg Leu Gly Gln Glu Arg Val Phe Trp Ile Cys Ile Ser
305 310 315 320

Leu Leu Glu Met Ala Tyr Ala Val Ala Ile Leu Val Gly Ser Thr Ser
325 330 335

Pro Tyr Leu Trp Ser Lys Val Ile Thr Val Ser Gly His Val Val Leu
340 345 350

Ala Ser Ile Leu Trp Gly Arg Ala Lys Ser Ile Asp Phe Lys Ser Lys
355 360 365

Ala Ala Leu Thr Ser Phe Tyr Met Phe Ile Trp Lys Leu Phe Tyr Ala
370 375 380

Glu Tyr Leu Leu Ile Pro Leu Val Arg
385 390

<210> 12
<211> 14
<212> PRT
<213> artificial sequence

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = w or y

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = k or r

<220>
<221> MISC_FEATURE

<222> (11)..(11)
<223> x = i or v

<400> 12

Ala Phe Xaa Xaa Phe Ser Arg Pro His Thr Xaa Ile Gly Thr
1 5 10

<210> 13
<211> 26
<212> PRT
<213> artificial sequence

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> x = v or i

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = e, w, f, or s

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> x = v or i

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> x = d or e

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> x = k or r

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> x = i or v

<220>
<221> MISC_FEATURE
<222> (22)..(22)
<223> x = h, n, t or y

<400> 13

Asn Xaa Tyr Ile Val Gly Leu Asn Gln Leu Xaa Asp Xaa Xaa Ile Asp
1 5 10 15

Xaa Xaa Asn Lys Pro Xaa Leu Pro Leu Ala
20 25

<210> 14
<211> 16
<212> PRT
<213> artificial sequence

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = i or l

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> x = i or v

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = i or m

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> x = r or k

<220>
<221> MISC_FEATURE
<222> (15)..(15)
<223> x = l, q, i, or v

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> x = y or f

<400> 14

Ile Ala Xaa Phe Lys Asp Xaa Pro Asp Xaa Glu Gly Asp Xaa Xaa Xaa
1 5 10 15

<210> 15
<211> 17

<212> PRT
<213> artificial sequence

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> x = f or c

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = q or m

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = f or y

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = i, l, or a

<220>
<221> MISC_FEATURE
<222> (15)..(15)
<223> x = i, m, or l

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> x = f, y, i, or l

<400> 15

Xaa Tyr Xaa Phe Ile Trp Lys Leu Phe Xaa Xaa Glu Tyr Leu Xaa Xaa
1 5 10 15

Pro

<210> 16
<211> 56
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 16

cgcgatttaa atggcgcgcc ctgcaggcgg ccgcctgcag ggcgcgccat ttaaat 56

<210> 17
<211> 32
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 17
tcgaggatcc gcggccgcaa gcttcctgca gg 32

<210> 18
<211> 32
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 18
tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 19
<211> 32
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 19
tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 20
<211> 32
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 20
tcgaggatcc gcggccgcaa gcttcctgca gg 32

<210> 21
<211> 36
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 21
tcgaggatcc gcggccgcaa gcttcctgca ggagct 36

<210> 22
<211> 28
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 22
cctgcaggaa gcttgcggcc gcggatcc 28

<210> 23
<211> 36
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 23
tcgacctgca ggaagcttgc ggccgcggat ccagct 36

<210> 24
<211> 28
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 24
ggatccgcgg ccgcaagctt cctgcagg 28

<210> 25
<211> 39
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 25
gatcacctgc aggaagcttg cggccgcgga tccaatgca 39

<210> 26
<211> 31
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic Primer Sequence

<400> 26
ttggatccgc ggccgcaagc ttcctgcagg t 31

<210> 27
<211> 969
<212> DNA
<213> Nostoc punctiforme

<400> 27
atgagccaga gttctcaaaa cagccctttg ccacgcaaac ctgttcaatc atatttccat 60

tggttatacg ctttctggaa attctctcgc cctcacacga ttattggtac aagtctgagt 120

gtgttgagtt tgtatttaat tgctattgcc attagtaata ataccgcttc tttattcact 180

actcccggct ccctaagccc tctcttcggc gcatggattg cttgtctatg tggcaatgtt 240

tacattgtag ggctgaatca attagaagat gttgatattg acaagattaa taaacctcat 300

ttaccgttgg catcaggtga gttttctcaa cagacgggac aattaattgt tgcatctact 360

gggattttgg cactagttat ggcgtggcta actgggccat tcttgtttgg catggtaaca 420

attagtttgg ccattggtac tgcttattct ttaccgccaa ttcgcttaaa acagtttccc 480

ttttgggcag cgctgtgtat tttttcggta cgcggcacga ttgttaattt aggattgtat 540

ttgcactata gttgggcgct gaaacaaagc caaacaattc cgcctgtggt gtgggtgctg 600

acattgttta ttttggtgtt tacctttgcg atcgcaatct ttaaagatat cccagatata 660

gaaggcgatc gcctctacaa tattactact ttcacgatta aactagggtc ccaagctgtg 720

tttaatctag ctctttgggt gataactgtc tgttatctag ggataattct ggtaggagtg 780

ctacgcatcg cttcagttaa ccccatttt ctgataactg ctcatttggc gctgttggtt 840

tggatgtggt ggcggagttt ggcggtagac ttacaagata aaagtgcgat cgctcaattc 900

taccaattta tctggaaact ctttttttata gaatatctaa tttttcctat cgcctgcttt 960

ttggcttag 969

<210> 28
<211> 957
<212> DNA
<213> Anabaena sp.

<400> 28
atgaaccaaa gttcccaaga cagaccgttg cgacctaaac cattgcaatc atcttttcag 60

tggctttatg ctttttggaa attttccccgc ccacacacaa ttattggcac aagtctcagt 120

gttttgggct tatatttaat ttctatcgcc gtcagttcca ccggttttgc cctgacgcag 180

ataaactccg ttttaggagc atggctggcc tgtctctgtg gcaatgttta tattgtgggg 240

ttaaatcaat tagaagatat tgaaattgat aaagttaata aacctcattt acctctagct 300

tcgggagaat ttagccgcaa acaaggacgg ataattgtaa ttctcacggg aattaccgcc 360

atagtattag cttggttaaa tggcccttat ttatttggta tggtggcggt gagtttagcc 420

attggtactg cctattcttt accaccaatt cgtttaaaac agtttccctt ttgggcggcc 480

ttgtgtattt tttcagtaag gggaacgatt gttaatttag gattatatct gcacttcagt 540

tggctactac agaataaaca gtcaattcct ctacctgtat ggatattaac ggtatttatt 600

ttaatattta cctttgcgat cgccatcttt aaagatatcc ctgatatgga aggcgatcgc 660

ctctacaata ttaccactct caccatccaa ctagggccac aagctgtctt taatttggca 720

atgtgggtat taacggtttg ctacttgggt atggtgataa ttggtgtgct gcggctaggt 780

acaattaact cagtgtttct ggtcgtgact catttagtaa ttctctgttg gatgtggatg 840

cagagtttag ccgtagacat acatgacaaa acggcgatcg ctcaattcta tcaatttatt 900

tggaagctct ttttcctaga atatttaatg tttcccattg cctgtctttt agcttaa 957

<210> 29
<211> 927
<212> DNA
<213> Synechocystis sp.

<400> 29
atggcaacta tccaagcttt ttggcgcttc tcccgccccc ataccatcat tggtacaact 60

ctgagcgtct gggctgtgta tctgttaact attctcgggg atggaaactc agttaactcc 120

cctgcttccc tggatttagt gttcggcgct tggctggcct gcctgttggg taatgtgtac 180

attgtcggcc tcaaccaatt gtgggatgtg gacattgacc gcatcaataa gccgaatttg 240

cccctagcta acggagattt ttctatcgcc cagggccgtt ggattgtggg actttgtggc 300

gttgcttcct tggcgatcgc ctggggatta gggctatggc tggggctaac ggtgggcatt 360

agtttgatta ttggcacggc ctattcggtg ccgccagtga ggttaaagcg cttttccctg 420

ctggcggccc tgtgtattct gacggtgcgg ggaattgtgg ttaacttggg cttatttta 480

ttttttagaa ttggtttagg ttatcccccc actttaataa cccccatctg ggttttgact 540

ttatttatct tagttttcac cgtggcgatc gccatttta aagatgtgcc agatatggaa 600

ggcgatcggc aatttaagat tcaaacttta actttgcaaa tcggcaaaca aaacgttttt 660

cggggaacct taattttact cactggttgt tatttagcca tggcaatctg gggcttatgg 720

gcggctatgc ctttaaatac tgctttcttg attgtttccc atttgtgctt attagcctta 780

ctctggtggc ggagtcgaga tgtacactta gaaagcaaaa ccgaaattgc tagtttttat 840

cagtttattt ggaagctatt tttcttagag tacttgctgt atcccttggc tctgtggtta 900

cctaattttt ctaatactat tttttag 927

<210> 30
<211> 1569

<212> DNA
<213> Zea mays

<400> 30
ccacgcgtcc gcccggccaa gggatggacg cgcttcgcct acggccgtcc ctcctctccg 60

tgcggcccgg cgcggcccgc ccgcgagatc attttctacc accatgttgt tccatacaac 120

gaaatggtga aggacgaatt tgcttttcta gccaaaggac ccaaggtcct accttgcatc 180

accatcagaa attcttcgaa tggaaatcct cctattgtag gatatcacat cgatcattaa 240

atacttctgt taatgcttcg gggcaacagc tgcagtctga acctgaaaca catgattcta 300

caaccatctg gagggcaata tcatcttctc tagatgcatt ttacagattt tcccggccac 360

atactgtcat aggaacagca ttaagcatag tctcagtttc ccttctagct gtccagagct 420

tgtctgatat atcacctttg ttcctcactg gtttgctgga ggcagtggta gctgcccttt 480

tcatgaatat ctatattgtt ggactgaacc agttattcga cattgagata gacaaggtta 540

acaagccaac tcttccattg gcatctgggg aatacacccc tgcaactggg gttgcaatag 600

tttcggtctt tgccgctatg agctttggcc ttggatgggc tgttggatca caacctctgt 660

tttgggctct tttcataagc tttgttcttg ggactgcata ttcaatcaat ctgccgtacc 720

ttcgatggaa gagatttgct gttgttgcag cactgtgcat attagcagtc cgtgcagtga 780

ttgttcagct ggccttttct ctccacattc agacttttgt tttcaggaga ccggcagtgt 840

tttctaggcc attattattt gcaactggat ttatgacgtt cttctctgtt gtaatagcac 900

tattcaagga tatacctgac atcgaaggag accgcatatt cgggatccga tccttcagcg 960

tccggttagg gcaaaagaag gtcttttgga tctgcgttgg cttgcttgag atggcctaca 1020

gcgttgcgat actgatggga gctacctctt cctgtttgtg gagcaaaaca gcaaccatcg 1080

ctggccattc catacttgcc gcgatcctat ggagctgcgc gcgatcggtg gacctgacga 1140

gcaaagccgc aataacgtcc ttctacatgt tcatctggaa gctgttctac gcggagtacc 1200

tgctcatccc tctggtgcgg tgagcgcgag gcgaggtggt ggcagacgga tcggcgtcgg 1260

cggggcggca aacaactcca cgggagaact tgagtgccgg aagtaaactc ccgtttgaaa 1320

gttgaagcgt gcaccaccgg caccgggcag agagagacac ggtggctgga tggatacgga 1380

tggccccccc aataaattcc cccgtgcatg gtaccccacg ctgcttgatg atatcccatg 1440

tgtccgggtg atcgtctcta gagagattgg ttgcacaacg tccaacatag cccgtaggta 1500

ttgctaccac tgctagtatg atactccttc ctagtccttg ccaaaaaaaa aaaaaaaaaa 1560

aaaaaaaag 1569

<210> 31
<211> 1236

<212> DNA
<213> Glycine max (ppt2)

<400> 31
atggattcac tgcttcttcg atctttccct aatattaata acgcctcttc tctcaccacc 60

actggtgcaa atttctccag gactaaatct ttcgccaaca tttaccatgc aagttcttat 120

gtgccaaatg cttcatggca caataggaaa atccaaaaag aatataattt tttgaggttt 180

cggtggccaa gtttgaacca tcattacaaa ggcattgagg gagcgtgtac atgtaaaaaa 240

tgtaatataa aatttgttgt gaaagcgacc tctgaaaaat ctcttgagtc tgaacctcaa 300

gcttttgatc caaaaagcat tttggactct gtcaagaatt ccttggatgc tttctacagg 360

ttttccaggc ctcacacagt tattggcaca gcattaagca taatttctgt gtctcttctt 420

gctgttgaga aaatatcaga tatatctcca ttatttttta ctggtgtgtt ggaggctgtg 480

gttgctgccc tgtttatgaa tatttatatt gttggtttga atcaattgtc tgatgttgaa 540

atagacaaga taaacaagcc gtatcttcca ttagcatctg gggaatattc ctttgaaact 600

ggtgtcacta ttgttgcatc tttttcaatt ctgagttttt ggcttggctg ggttgtaggt 660

tcatggccat tattttgggc ccttttttgta agctttgtgc taggaactgc ttattcaatc 720

aatgtgcctc tgttgagatg gaagaggttt gcagtgcttg cagcgatgtg cattctagct 780

gttcgggcag taatagttca acttgcattt ttccttcaca tgcagactca tgtgtacaag 840

aggccacctg tcttttcaag accattgatt tttgctactg cattcatgag cttcttctct 900

gtagttatag cactgtttaa ggatatacct gacattgaag gagataaagt atttggcatc 960

caatcttttt cagtgcgttt aggtcagaag ccggtgttct ggacttgtgt tacccttctt 1020

gaaatagctt atggagtcgc cctcctggtg ggagctgcat ctccttgtct ttggagcaaa 1080

attttcacgg gtctgggaca cgctgtgctg gcttcaattc tctggtttca tgccaaatct 1140

gtagatttga aaagcaaagc ttcgataaca tccttctata tgtttatttg gaagctattt 1200

tatgcagaat acttactcat tccttttgtt agatga 1236

<210> 32
<211> 1188
<212> DNA
<213> Glycine max (ppt1)

<400> 32
atggattcga tgcttcttcg atcttttcct aatattaaca acgcttcttc tctcgccacc 60

actggttctt atttgccaaa tgcttcatgg cacaatagga aaatccaaaa agaatataat 120

tttttgaggt ttcggtggcc aagtttgaac caccattaca aaagcattga aggagggtgt 180

acatgtaaaa aatgtaatat aaaatttgtt gtgaaagcga cctctgaaaa atcttttgag 240

tctgaacccc aagcttttga tccaaaaagc attttggact ctgtcaagaa ttccttggat 300

gctttctaca ggttttccag acctcacaca gttattggca cagcattaag cataatttct 360

gtgtccctcc ttgctgttga gaaaatatca gatatatctc cattattttt tactggtgtg 420

ttggaggctg tggttgctgc cctgtttatg aatatttata ttgttggttt gaatcaattg 480

tctgatgttg aaatagacaa gataaacaag ccgtatcttc cattagcatc tggggaatat 540

tcctttgaaa ctggtgtcac tattgttgca tctttttcaa ttctgagttt ttggcttggc 600

tgggttgtag gttcatggcc attatttttgg gccctttttg taagctttgt gctaggaact 660

gcttattcaa tcaatgtgcc tctgttgaga tggaagaggt ttgcagtgct tgcagcgatg 720

tgcattctag ctgttcgggc agtaatagtt caacttgcat ttttccttca catccagact 780

catgtataca agaggccacc tgtcttttca agatcattga tttttgctac tgcattcatg 840

agcttcttct ctgtagttat agcactgttt aaggatatac ctgacattga aggagataaa 900

gtatttggca tccaatcttt ttcagtgcgt ttaggtcaga agccggtatt ctggacttgt 960

gttatccttc ttgaaatagc ttatggagtc gccctcctgg tgggagctgc atctccttgt 1020

ctttggagca aaattgtcac gggtctggga cacgctgttc tggcttcaat tctctggttt 1080

catgccaaat ctgtagattt gaaaagcaaa gcttcgataa catccttcta tatgtttatt 1140

tggaagctat tttatgcaga atacttactc attccttttg ttagatga 1188

<210> 33
<211> 1182
<212> DNA
<213> Arabidopsis thaliana

<400> 33
atggagtctc tgctctctag ttcttctctt gtttccgctg ctggtgggtt ttgttggaag 60

aagcagaatc taaagctcca ctctttatca gaaatccgag ttctgcgttg tgattcgagt 120

aaagttgtcg caaaaccgaa gtttaggaac aatcttgtta ggcctgatgg tcaaggatct 180

tcattgttgt tgtatccaaa acataagtcg agatttcggg ttaatgccac tgcgggtcag 240

cctgaggctt tcgactcgaa tagcaaacag aagtctttta gagactcgtt agatgcgttt 300

tacaggtttt ctaggcctca tacagttatt ggcacagtgc ttagcatttt atctgtatct 360

ttcttagcag tagagaaggt ttctgatata tctcctttac ttttcactgg catcttggag 420

gctgttgttg cagctctcat gatgaacatt tacatagttg ggctaaatca gttgtctgat 480

gttgaaatag ataaggttaa caagccctat cttccattgg catcaggaga atattctgtt 540

aacaccggca ttgcaatagt agcttccttc tccatcatga gtttctggct tgggtggatt 600

gttggttcat ggccattgtt ctgggctctt tttgtgagtt tcatgctcgg tactgcatac 660

tctatcaatt tgccactttt acggtggaaa agatttgcat tggttgcagc aatgtgtatc 720

ctcgctgtcc gagctattat tgttcaaatc gccttttatc tacatattca gacacatgtg 780

tttggaagac caatcttgtt cactaggcct cttattttcg ccactgcgtt tatgagcttt 840

ttctctgtcg ttattgcatt gtttaaggat atacctgata tcgaagggga taagatattc 900

ggaatccgat cattctctgt aactctgggt cagaaacggg tgttttggac atgtgttaca 960

ctacttcaaa tggcttacgc tgttgcaatt ctagttggag ccacatctcc attcatatgg 1020

agcaaagtca tctcggttgt gggtcatgtt atactcgcaa caactttgtg ggctcgagct 1080

aagtccgttg atctgagtag caaaaccgaa ataacttcat gttatatgtt catatggaag 1140

ctcttttatg cagagtactt gctgttacct tttttgaagt ga 1182

<210> 34
<211> 1374
<212> DNA
<213> Cuphea pulcherrima

<400> 34
ccacgcgtcc ggctggtttg tgggttttgc gagcacgagg aaggaaaaaa catgcggatg 60

gagtctctgc ttctgaattc tttctctcca tctccggcgg gaggaaaaat ttgtagggcc 120

gatacttaca agaaggccta cttcgcaact gcgaggtgca acacattgaa cagcctcaac 180

aagaatacag gtgaatatca tctcagcaga acccgacaac ggttcacatt tcaccaaaat 240

ggtcacagaa cttacctagt caaggcagtg tccgggcagt ccctggagtc tgagcccgaa 300

agttacccta acaataggtg ggattatgtc aaaagtgctg ctgatgcctt ctaccggttt 360

tctcgtcccc acacaattat aggcactgcg ttgagcatag tatcggtttc gcttcttgct 420

gtagagaagt tgcctgaatt gaattcaatg tttttcactg gcttattgga ggtgattttg 480

gctgccctct tcatgaatat atatattgtc ggtttgaatc agttgtctga tatagacatt 540

gacaaggtaa acaagccgta tcttcccctg gcatcaggag aattctcggt tggaactggg 600

gttaccattg taacatcctt cttgattatg agcttttggc tggggtgggt tgtcggttca 660

tggcccttgt tttgggccct tttcatcagt tttgtgcttg gaacagcata ctcaatcgat 720

atgccaatgc tcagatggaa gagatctgca gttgtggctg cactgtgcat tctagctgtt 780

cgggccgtga ttgttcagat agcgtttttt ttgcacatgc agatgcatgt gtatggaaga 840

gcagctgcac tttctcggcc tgtaatattt gccacaggct ttatgagctt cttttctatt 900

gttattgcgt tgtttaagga cattcctgac atagaaggtg ataaaatatt tgggatccgg 960

tcattcactg ttcgtctggg ccaagaacgg gttttctgga tatgcatatc acttctcgaa 1020

atggcttatg ctgttgcgat tcttgttggg tcgacgtctc cctatctttg gagcaaagtc 1080

atcacggttt cgggtcatgt tgtgttggcc tccatactat ggggacgagc caagtctatc 1140

gactttaaga gcaaagcagc actaacctcc ttctacatgt ttatttggaa gctattttac 1200

gcagaatact tgcttatacc gcttgtacga tgagctttcg ggatcagaac attacattat 1260

cgtaaactga acaatttaga attgcatatt gttcagatga cagctccatc ttggcaataa 1320

aatttgatat gaatgtctct gatccaaaaa aaaaaaaaaa aaaaaaaaaa aaag 1374

<210> 35
<211> 1486
<212> DNA
<213> Allium porrum

<400> 35
gcacgagttt tgaagaatgt taagcatgga ctccctcctt accaagccag ttgtaatacc 60

tctgccttct ccagtttgtt cactaccaat cttgcgaggc agttctgcac cagggcagta 120

ttcatgtaga aactacaatc caataagaat tcaaaggtgc ctcgtaaatt atgaacatgt 180

gaaaccaagg tttacaacat gtagtaggtc tcaaaaactt ggtcatgtaa aagccacatc 240

cgagcattct ttagaatctg gatccgaagg atacactcct agaagcatat gggaagccgt 300

actagcttca ctgaatgttc tatacaaatt ttcacgacct cacacaataa taggaacagc 360

aatgggcata atgtcagttt ctttgcttgt tgtcgagagc ctatccgata tttctcctct 420

gtttttttgtg ggattattag aggctgtggt tgctgcattg tttatgaatg tttacattgt 480

aggtctgaat caattatttg acatagaaat agacaaggtc aataaacctg atcttcctct 540

tgcatctgga gaatactcac caagagctgg tactgctatt gtcattgctt cagccatcat 600

gagctttggc attggatggt tagttggctc ttggccatta ttctgggcgc tttttattag 660

tttcgttctt ggcactgcat attcaatcaa tctaccattt ctaagatgga agagatccgc 720

cgttgttgca gcaatatgta tccttgctgt acgagcagtt atagtccagc tcgccttttt 780

cttacacata cagagttttg ttttcaaaag accagcaagt ttcacaaggc cattgatatt 840

tgcaactgcc ttcatgagct tcttctcagt tgttattgct ctatttaagg atatacctga 900

tatagacgga gacaaaatat ttggcatcca ttctttcagc gtgcgccttg gccaggagag 960

ggtgttttgg atatgtatat atctccttga gatggcctac actgttgtca tggttgttgg 1020

agctacttcc tcatgcctat ggagcaaatg cttaacagtg ataggtcatg caattcttgg 1080

gtcgttactt tggaatcgtg ctagatctca tggaccaatg accaaaacca ctattacatc 1140

tttttatatg ttcgtgtgga agctcttcta tgctgagtac ttgctcattc cttttgtaag 1200

atgagggttt atgacctaca tggaaaagaa tcgcaagaga agatgagtag ataatggagg 1260

cagaaatggc tggaattaac aacgctttaa ttgtcatctt aaaaacggag agttctttca 1320

acaattgcag atcatttctc cttaattata ttcatgttgt atgttgtgtt aaagattatc 1380

attgaatgac aatagcctat gttgaattta ggatatccag tggttttctt tgttcttttt 1440

taagaattta ttcacagaaa aatgaagtaa aaaaaaaaaa aaaaaa 1486

<210> 36
<211> 1670
<212> DNA
<213> Triticum sp.

<400> 36
ccacgcgtcc ggtcccactg cccgcccccc acccgcgcgc cgccgcggcg atggactcgc 60

tccgcctccg gccgtcctcg ctccgctccg cgcccggcgc cgccgccgcc cgccggcgag 120

atcatattct accatcattt tgttcgatcc aacgaaatgg taaagggcga gttactttgt 180

ccatccaagc atccaaaggc cctaccatta atcactgtaa aaagttcttg gattggaaat 240

attccaacca taggatatca catcaatcaa taaatacttc tgcaaaagct gggcaatcgc 300

tacagcctga aactgaagca cacgatcctg caagcttctg gaagccaata tcatcttctc 360

tggacgcgtt ctacaggttt tctcggccac ataccatcat aggaacagca ctaagcatag 420

tctcagtttc ccttctagct gtcgagagct tatctgatat ttcgcccttg ttcctcactg 480

gtttgctgga ggcagtggtg gctgctcttt ttatgaacat ctatattgtt ggattgaatc 540

agttgttcga cattgaaatt gacaaggtta acaagccaac tcttccacta gcatctgggg 600

aatactctcc tgcaactgga gttgcaatag tgtcagtatt tgcagccatg agctttggcc 660

ttggatgggt tgttggatca ccacctctgt tttgggctct ttttattagc tttgttcttg 720

gaactgctta ttcagtcaat ctgccgtact ttcgatggaa gagatctgct gttgttgcag 780

cactctgcat attagcagtg cgtgcggtga ttgttcaact ggcatttttt ctccacattc 840

agacatttgt tttcagaagg ccggcagtct tttcaaagcc attgatattt gcaactgcct 900

tcatgacatt cttctcagtt gtaatagcat tattcaagga tatacccgat attgaagggg 960

accgcatctt tggaattcaa tcttttagtg ttagattagg tcaaagcaag gttttctgga 1020

cttgtgttgg tctacttgag gttgcctacg gtgttgcgat actgatgggg gtaacttctt 1080

ccagtttgtg gagcaaatct ctaactgttg tgggccatgc aatcctcgcc agcatcttgt 1140

ggagcagcgc acggtccatc gacttgacaa gcaaagctgc gataacatcc ttctacatgc 1200

tcatctggag gctgttctac gcggagtacc tgctcatccc tctggtgaga tgaggaccga 1260

caagcagccc acggaagaac ttcagtgccg gagtacagct gtgcgaatcc atttgaattt 1320

cggatggtca cggaccgcgc ccaataaaac tcccagagcc ttgccccggt acatcgttga 1380

ttttccagcc atgaatggtg agatcaccac ctaaagatgg ataacctccc catgtaccca 1440

agctgggcca ggtgagctgt agtttagttg atgctagcga gcaacaactc ctgcagcagg 1500

cacgcggctg cctggaaaat aaggctcccc actcccaatt acattctgtt gtacggtttt 1560

agtacttgtg aattttgctc tggtccgttg ttgtctagga tgtttggaac attgcgcaga 1620

ctttcttata tcttaccggg aggggtgaat tggcaaaaaa aaaaaaaaag 1670

<210> 37
<211> 41
<212> DNA
<213> artificial sequence

<220>
<223> Primer Sequence

<400> 37
ggatccgcgg ccgcacaatg gagtctctgc tctctagttc t 41

<210> 38
<211> 38
<212> DNA
<213> artificial sequence

<220>
<223> Primer Sequence

<400> 38
ggatcctgca ggtcacttca aaaaaggtaa cagcaagt 38

<210> 39
<211> 14
<212> PRT
<213> artificial sequence

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> x = a or v

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> x = f or l

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = w or y

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = k or r

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = i or v

<400> 39

Xaa Xaa Xaa Xaa Phe Ser Arg Pro His Thr Xaa Ile Gly Thr
1 5 10

<210> 40
<211> 26
<212> PRT
<213> artificial sequence

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> x = i or l

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = w, e, s, or f

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> x = v or i

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> x = d or e

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> x = r or k

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> x = i or v

<220>
<221> MISC_FEATURE
<222> (23)..(23)
<223> x = n, h, y, t, or d

<400> 40

Asn Xaa Tyr Ile Val Gly Leu Asn Gln Leu Xaa Asp Xaa Xaa Ile Asp
1 5 10 15

Xaa Xaa Asn Lys Pro Xaa Leu Pro Leu Ala
20 25

<210> 41
<211> 16
<212> PRT
<213> artificial sequence

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = i or l

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> x = v or i

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = m or i

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = e or d

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> x = r or k

<220>
<221> MISC_FEATURE
<222> (15)..(15)
<223> x = i, v, l, or q

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> x = y or f

<400> 41

Ile Ala Xaa Phe Lys Asp Xaa Pro Asp Xaa Xaa Gly Asp Xaa Xaa Xaa
1 5 10 15

<210> 42
<211> 17
<212> PRT
<213> artificial sequence

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> x = f or c

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = m or q

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = f or l

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> x = i or v

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> x = k or r

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = f or y

<220>
<221> MISC_FEATURE

<222> (11)..(11)
<223> x = l, i or a

<220>
<221> MISC_FEATURE
<222> (15)..(15)
<223> x = l, m, or i

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> x = f, y, l, or i

<400> 42

Xaa Tyr Xaa Xaa Xaa Trp Xaa Leu Phe Xaa Xaa Glu Tyr Leu Xaa Xaa
1 5 10 15

Pro

<210> 43
<211> 349
<212> PRT
<213> Trichodesmium erythraeum

<400> 43

Met Gly Lys Ile Ala Gly Ser Gln Gln Gly Lys Ile Thr Thr Asn Trp
1 5 10 15

Leu Gln Lys Tyr Val Pro Trp Leu Tyr Ser Phe Trp Lys Phe Ala Arg
20 25 30

Pro His Thr Ile Ile Gly Thr Ser Leu Ser Val Leu Ala Leu Tyr Ile
35 40 45

Ile Ala Met Gly Asp Arg Ser Asn Phe Phe Asp Lys Tyr Phe Phe Leu
50 55 60

Tyr Ser Leu Ile Leu Leu Leu Ile Thr Trp Ile Ser Cys Leu Cys Gly
65 70 75 80

Asn Ile Tyr Ile Val Gly Leu Asn Gln Leu Glu Asp Ile Glu Ile Asp
85 90 95

Arg Ile Asn Lys Pro His Leu Pro Ile Ala Ala Gly Glu Phe Ser Arg
100 105 110

Phe Ser Gly Gln Ile Ile Val Val Ile Thr Gly Ile Leu Ala Leu Ser
115 120 125

Phe Ala Gly Leu Gly Gly Pro Phe Leu Leu Gly Thr Val Gly Ile Ser
130 135 140

Leu Ala Ile Gly Thr Ala Tyr Ser Leu Pro Pro Ile Arg Leu Lys Arg
145 150 155 160

Phe Pro Val Leu Ala Ala Leu Cys Ile Phe Thr Val Arg Gly Val Ile
165 170 175

Val Asn Leu Gly Ile Phe Leu Ser Phe Val Trp Gly Phe Glu Lys Val
180 185 190

Glu Glu Val Ser Gly Gly Leu Ile Lys Trp Met Gly Glu Leu Gly Glu
195 200 205

Val Val Leu Leu Gln Lys Ser Leu Met Val Pro Glu Ile Pro Leu Thr
210 215 220

Val Trp Ala Leu Thr Leu Phe Val Ile Val Phe Thr Phe Ala Ile Ala
225 230 235 240

Ile Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Arg Gln Tyr Asn Ile
245 250 255

Asn Thr Phe Thr Ile Lys Leu Gly Ala Phe Ala Val Phe Asn Leu Ala
260 265 270

Arg Trp Val Leu Thr Phe Cys Tyr Leu Gly Met Val Met Val Gly Val
275 280 285

Val Trp Leu Ala Ser Val Asn Leu Phe Phe Leu Val Ile Ser His Leu
290 295 300

Leu Ala Leu Gly Ile Met Trp Trp Phe Ser Gln Arg Val Asp Leu His
305 310 315 320

Asp Lys Lys Ala Ile Ala Asp Phe Tyr Gln Phe Ile Trp Lys Leu Phe
325 330 335

Phe Leu Glu Tyr Leu Ile Phe Pro Met Ala Cys Phe Phe
340 345

<210> 44

<211> 300
<212> PRT
<213> Chloroflexus aurantiacus

<400> 44

Met Arg Lys Gln Leu Arg Leu Leu Ile Glu Phe Ala Arg Pro His Thr
1 5 10 15

Val Ile Ala Thr Ser Val Gln Val Leu Thr Met Leu Ile Ile Val Ile
20 25 30

Gly Trp His Pro Pro Thr Leu Glu Leu Val Gly Leu Val Gly Val Thr
35 40 45

Leu Val Val Cys Leu Ala Leu Asn Leu Tyr Val Val Gly Val Asn Gln
50 55 60

Leu Thr Asp Val Ala Ile Asp Arg Ile Asn Lys Pro Trp Leu Pro Val
65 70 75 80

Ala Ala Gly Gln Leu Ser Ser Asp Ala Ala Gln Arg Ile Val Ile Ser
85 90 95

Ala Leu Phe Ile Ala Leu Thr Gly Ala Ala Met Leu Gly Pro Pro Leu
100 105 110

Trp Trp Thr Val Ser Ile Ile Ala Leu Ile Gly Ser Leu Tyr Ser Leu
115 120 125

Pro Pro Leu Arg Leu Lys Arg His Pro Leu Ala Ala Ala Leu Ser Ile
130 135 140

Ala Gly Ala Arg Gly Val Ile Ala Asn Leu Gly Leu Ala Phe His Tyr
145 150 155 160

Gln Tyr Trp Leu Asp Ser Glu Leu Pro Ile Thr Thr Leu Ile Leu Val
165 170 175

Ala Thr Phe Phe Phe Gly Phe Ala Met Val Ile Ala Leu Tyr Lys Asp
180 185 190

Leu Pro Asp Asp Arg Gly Asp Arg Leu Tyr Gln Ile Glu Thr Leu Thr
195 200 205

Thr Arg Leu Gly Pro Gln Arg Val Leu His Leu Gly Arg Ile Leu Leu
210 215 220

Thr Ala Cys Tyr Leu Leu Pro Ile Ala Val Gly Leu Trp Ser Leu Pro
225 230 235 240

Thr Phe Ala Ala Ala Phe Leu Ala Leu Ser His Val Val Val Ile Ser
245 250 255

Val Phe Trp Leu Val Ser Met Arg Val Asp Leu Gln Arg Arg Gln Ser
260 265 270

Ile Ala Ser Phe Tyr Met Phe Leu Trp Gly Ile Phe Tyr Thr Glu Phe
275 280 285

Ala Leu Leu Ser Ile Tyr Arg Leu Thr Tyr Thr Leu
290 295 300

<210> 45
<211> 402
<212> PRT
<213> Arabidopsis thaliana

<400> 45

Met Glu Leu Ser Ile Ser Gln Ser Pro Arg Val Arg Phe Ser Ser Leu
1 5 10 15

Ala Pro Arg Phe Leu Ala Ala Ser His His His Arg Pro Ser Val His
20 25 30

Leu Ala Gly Lys Phe Ile Ser Leu Pro Arg Asp Val Arg Phe Thr Ser
35 40 45

Leu Ser Thr Ser Arg Met Arg Ser Lys Phe Val Ser Thr Asn Tyr Arg
50 55 60

Lys Ile Ser Ile Arg Ser Val Cys Ala Phe Cys Asn Gly Thr His Lys
65 70 75 80

Ser Arg Tyr Tyr Gln Ala Cys Ser Gln Val Gly Ala Ala Glu Ser Asp
85 90 95

Asp Pro Val Leu Asp Arg Ile Ala Arg Phe Gln Asn Ala Cys Trp Arg
100 105 110

Phe Leu Arg Pro His Thr Ile Arg Gly Thr Ala Leu Gly Ser Thr Ala
115 120 125

Leu Val Thr Arg Ala Leu Ile Glu Asn Thr His Leu Ile Lys Trp Ser

130 135 140

Leu Val Leu Lys Ala Leu Ser Gly Leu Leu Ala Leu Ile Cys Gly Asn
145 150 155 160

Gly Tyr Ile Val Gly Ile Asn Gln Ile Tyr Asp Ile Gly Ile Asp Lys
165 170 175

Val Asn Lys Pro Tyr Leu Pro Ile Ala Ala Gly Asp Leu Ser Val Gln
180 185 190

Ser Ala Trp Leu Leu Val Ile Phe Phe Ala Ile Ala Gly Leu Leu Val
195 200 205

Val Gly Phe Asn Phe Gly Pro Phe Ile Thr Ser Leu Tyr Ser Leu Gly
210 215 220

Leu Phe Leu Gly Thr Ile Tyr Ser Val Pro Pro Leu Arg Met Lys Arg
225 230 235 240

Phe Pro Val Ala Ala Phe Leu Ile Ile Ala Thr Val Arg Gly Phe Leu
245 250 255

Leu Asn Phe Gly Val Tyr His Ala Thr Arg Ala Ala Leu Gly Leu Pro
260 265 270

Phe Gln Trp Ser Ala Pro Val Ala Phe Ile Thr Ser Phe Val Thr Leu
275 280 285

Phe Ala Leu Val Ile Ala Ile Thr Lys Asp Leu Pro Asp Val Glu Gly
290 295 300

Asp Arg Lys Phe Gln Ile Ser Thr Leu Ala Thr Lys Leu Gly Val Arg
305 310 315 320

Asn Ile Ala Phe Leu Gly Ser Gly Leu Leu Leu Val Asn Tyr Val Ser
325 330 335

Ala Ile Ser Leu Ala Phe Tyr Met Pro Gln Val Phe Arg Gly Ser Leu
340 345 350

Met Ile Pro Ala His Val Ile Leu Ala Ser Gly Leu Ile Phe Gln Thr
355 360 365

Trp Val Leu Glu Lys Ala Asn Tyr Thr Lys Glu Ala Ile Ser Gly Tyr
370 375 380

Tyr Arg Phe Ile Trp Asn Leu Phe Tyr Ala Glu Tyr Leu Leu Phe Pro
385 390 395 400

Phe Leu

<210> 46
<211> 12
<212> PRT
<213> Artificial

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> x = w or y

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> x = r or k

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = l or s

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> x = i or v

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = i or r

<400> 46

Xaa Xaa Phe Xaa Arg Pro His Thr Xaa Xaa Gly Thr
1 5 10

<210> 47
<211> 26
<212> PRT
<213> Artificial

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> x = v, i, or g

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> x = i or l

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = i or l

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = s, f, y, or e

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> x = v or i

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> x = r, s, g, e, or d

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> x = k or r

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> x = v or i

<220>
<221> MISC_FEATURE
<222> (22)..(22)
<223> x = y, d, t, n, or h

<220>
<221> MISC_FEATURE
<222> (25)..(25)
<223> x = i or l

<400> 47

Asn Xaa Tyr Ile Val Gly Xaa Asn Gln Xaa Xaa Asp Xaa Xaa Ile Asp
1 5 10 15

Xaa Xaa Asn Lys Pro Xaa Leu Pro Xaa Ala
20 25

<210> 48
<211> 14
<212> PRT
<213> Arabidopsis thaliana

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = l or i

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = f or t

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> x = l, i, or v

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = i, v, or m

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = e or d

<400> 48

Ile Ala Xaa Xaa Lys Asp Xaa Pro Asp Xaa Xaa Gly Asp Arg
1 5 10

<210> 49
<211> 23
<212> PRT
<213> Artificial

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE

<222> (2)..(2)
<223> x = d, e, t, a, or s

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = a, e, s, or t

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = i or l

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> x = s, t, or a

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> x = q, g, or s

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> x = f, y, or c

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> x = q, m, or r

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = f or l

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = i or v

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> x = n or k

<220>
<221> MISC_FEATURE
<222> (16)..(16)

<223> x = y or f

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> x = a, l, or i

<220>
<221> MISC_FEATURE
<222> (21)..(21)
<223> x = f, i, l, or m

<220>
<221> MISC_FEATURE
<222> (22)..(22)
<223> x = f, l, i, or y

<220>
<221> MISC_FEATURE
<222> (20)..(20)
<223> x = a, i, or l

<400> 49

Lys Xaa Xaa Xaa Xaa Xaa Xaa Tyr Xaa Xaa Xaa Trp Xaa Leu Phe Xaa
1 5 10 15

Xaa Glu Tyr Xaa Xaa Xaa Pro
20

<210> 50
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 50
gtggctcggc ttcacttttt ac 22

<210> 51
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 51
ccacactcat atcaccgtgg 20

<210> 52
<211> 27
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 52
cagtgctgga tagaattgcc cggttcc 27

<210> 53
<211> 26
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 53
gagatctatc agtgcagtct gcttgg 26

<210> 54
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 54
gggacaagca tttttattgc aag 23

<210> 55
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 55
gccaagatca catgtgcagg aatc 24

<210> 56
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 56
gccaagatca catgtgcagg aatc 24

<210> 57

<211> 386
<212> PRT
<213> Arabidopsis thaliana

<400> 57

Met Glu Leu Ser Ile Ser Gln Ser Pro Arg Val Arg Phe Ser Ser Leu
1 5 10 15

Ala Pro Arg Phe Leu Ala Ala Ser His His His Arg Pro Ser Val His
20 25 30

Leu Ala Gly Lys Phe Ile Ser Leu Pro Arg Asp Val Arg Phe Thr Ser
35 40 45

Leu Ser Thr Ser Arg Met Arg Ser Lys Phe Val Ser Thr Asn Tyr Arg
50 55 60

Lys Ile Ser Ile Arg Ala Cys Ser Gln Val Gly Ala Ala Glu Ser Asp
65 70 75 80

Asp Pro Val Leu Asp Arg Ile Ala Arg Phe Gln Asn Ala Cys Trp Arg
85 90 95

Phe Leu Arg Pro His Thr Ile Arg Gly Thr Ala Leu Gly Ser Thr Ala
100 105 110

Leu Val Thr Arg Ala Leu Ile Glu Asn Thr His Leu Ile Lys Trp Ser
115 120 125

Leu Val Leu Lys Ala Leu Ser Gly Leu Leu Ala Leu Ile Cys Gly Asn
130 135 140

Gly Tyr Ile Val Gly Ile Asn Gln Ile Tyr Asp Ile Gly Ile Asp Lys
145 150 155 160

Val Asn Lys Pro Tyr Leu Pro Ile Ala Ala Gly Asp Leu Ser Val Gln
165 170 175

Ser Ala Trp Leu Leu Val Ile Phe Phe Ala Ile Ala Gly Leu Leu Val
180 185 190

Val Gly Phe Asn Phe Gly Pro Phe Ile Thr Ser Leu Tyr Ser Leu Gly
195 200 205

Leu Phe Leu Gly Thr Ile Tyr Ser Val Pro Pro Leu Arg Met Lys Arg
210 215 220

Phe Pro Val Ala Ala Phe Leu Ile Ile Ala Thr Val Arg Gly Phe Leu
225 230 235 240

Leu Asn Phe Gly Val Tyr His Ala Thr Arg Ala Ala Leu Gly Leu Pro
245 250 255

Phe Gln Trp Ser Ala Pro Val Ala Phe Ile Thr Ser Phe Val Thr Leu
260 265 270

Phe Ala Leu Val Ile Ala Ile Thr Lys Asp Leu Pro Asp Val Glu Gly
275 280 285

Asp Arg Lys Phe Gln Ile Ser Thr Leu Ala Thr Lys Leu Gly Val Arg
290 295 300

Asn Ile Ala Phe Leu Gly Ser Gly Leu Leu Leu Val Asn Tyr Val Ser
305 310 315 320

Ala Ile Ser Leu Ala Phe Tyr Met Pro Gln Val Phe Arg Gly Ser Leu
325 330 335

Met Ile Pro Ala His Val Ile Leu Ala Ser Gly Leu Ile Phe Gln Thr
340 345 350

Trp Val Leu Glu Lys Ala Asn Tyr Thr Lys Glu Ala Ile Ser Gly Tyr
355 360 365

Tyr Arg Phe Ile Trp Asn Leu Phe Tyr Ala Glu Tyr Leu Leu Phe Pro
370 375 380

Phe Leu
385

<210> 58
<211> 379
<212> PRT
<213> Oryza sativa

<400> 58

Met Ala Ser Leu Ala Ser Pro Pro Leu Pro Cys Arg Ala Ala Ala Thr
1 5 10 15

Ala Ser Arg Ser Gly Arg Pro Ala Pro Arg Leu Leu Gly Pro Pro Pro
20 25 30

Pro Pro Ala Ser Pro Leu Leu Ser Ser Ala Ser Ala Arg Phe Pro Arg

35 40 45

Ala Pro Cys Asn Ala Ala Arg Trp Ser Arg Arg Asp Ala Val Arg Val
50 55 60

Cys Ser Gln Ala Gly Ala Ala Gly Pro Ala Pro Leu Ser Lys Thr Leu
65 70 75 80

Ser Asp Leu Lys Asp Ser Cys Trp Arg Phe Leu Arg Pro His Thr Ile
85 90 95

Arg Gly Thr Ala Leu Gly Ser Ile Ala Leu Val Ala Arg Ala Leu Ile
100 105 110

Glu Asn Pro Gln Leu Ile Asn Trp Trp Leu Val Phe Lys Ala Phe Tyr
115 120 125

Gly Leu Val Ala Leu Ile Cys Gly Asn Gly Tyr Ile Val Gly Ile Asn
130 135 140

Gln Ile Tyr Asp Ile Arg Ile Asp Lys Val Asn Lys Pro Tyr Leu Pro
145 150 155 160

Ile Ala Ala Gly Asp Leu Ser Val Gln Thr Ala Trp Leu Leu Val Val
165 170 175

Leu Phe Ala Ala Ala Gly Phe Ser Ile Val Val Thr Asn Phe Ile Leu
180 185 190

Phe Ile Thr Ser Leu Tyr Cys Leu Ala Leu Phe Leu Gly Thr Ile Tyr
195 200 205

Ser Val Pro Pro Phe Arg Leu Lys Arg Tyr Arg Ala Pro Ala Cys Leu
210 215 220

Ile Ile Ala Thr Val Arg Gly Phe Leu Arg Asn Leu Gly Val Tyr Tyr
225 230 235 240

Ala Thr Arg Ala Ala Leu Gly Leu Thr Phe Gln Trp Ser Ser Pro Val
245 250 255

Ala Phe Ile Thr Cys Phe Val Thr Leu Phe Ala Leu Val Ile Ala Ile
260 265 270

Thr Lys Asp Leu Pro Asp Val Glu Gly Asp Arg Lys Tyr Gln Ile Ser
275 280 285

Thr Leu Ala Thr Lys Leu Gly Val Arg Asn Ile Ala Phe Leu Gly Ser
290 295 300

Gly Leu Leu Ile Ala Asn Tyr Val Ala Ala Ile Ala Val Ala Phe Leu
305 310 315 320

Met Pro Gln Ala Phe Arg Arg Thr Val Met Val Pro Val His Ala Ala
325 330 335

Leu Ala Val Gly Ile Ile Phe Gln Thr Trp Val Leu Glu Gln Ala Lys
340 345 350

Tyr Thr Lys Asp Ala Ile Ser Gln Tyr Tyr Arg Phe Ile Trp Asn Leu
355 360 365

Phe Tyr Ala Glu Tyr Ile Phe Phe Pro Leu Ile
370 375

<210> 59
<211> 1423
<212> DNA
<213> Arabidopsis thaliana

<400> 59
ctctcactac agaacataca caagtataat tcgtcgatcg acccacgcgt ccggcagagc 60

aaagagtttt tgtgtggcta gtggcatcaa tggagctctc gatctcacaa tcaccgcgtg 120

ttcggttctc gtctctggcg cctcgtttct tagcagcttc tcatcatcat cgtccttctg 180

tgcatttagc tgggaagttt ataagcctcc ctcgagatgt tcgcttcacg agcttatcaa 240

cttcaagaat gcggtccaaa tttgtttcaa ccaattatag aaaaatctca atccgggcat 300

gttctcaggt tggtgctgct gagtctgatg atccagtgct ggatagaatt gcccggttcc 360

aaaatgcttg ctggagattt cttagacccc atacaatccg cggaacagct ttaggatcca 420

ctgccttggt gacaagagct ttgatagaga acactcattt gatcaaatgg agtcttgtac 480

taaaggcact ttcaggtctt cttgctctta tttgtgggaa tggttatata gtcggcatca 540

atcagatcta cgacattgga atcgacaaag tgaacaaacc atacttgcca atagcagcag 600

gagatctatc agtgcagtct gcttggttgt tagtgatatt ttttgcgata gcagggcttt 660

tagttgtcgg atttaacttt ggtccattca ttacaagcct atactctctt ggccttttc 720

tgggaaccat ctattctgtt ccacccctca gaatgaaaag attcccagtt gcagcatttc 780

ttattattgc cacggtacga ggtttccttc ttaactttgg tgtgtaccat gctacaagag 840

ctgctcttgg acttccattt cagtggagtg cacctgtggc gttcatcaca tcttttgtga 900

cactgtttgc actggtcatt gctattacaa aggaccttcc tgatgttgaa ggagatcgaa 960

agttccaaat atcaaccctg gcaacaaaac ttggagtgag aaacattgca ttcctcggtt 1020

ctggacttct gctagtaaat tatgtttcag ccatatcact agctttctac atgcctcagg 1080

tttttagagg tagcttgatg attcctgcac atgtgatctt ggcttcaggc ttaattttcc 1140

agacatgggt actagaaaaa gcaaactaca ccaaggaagc tatctcagga tattatcggt 1200

ttatatggaa tctcttctac gcagagtatc tgttattccc cttcctctag ctttcaattt 1260

catggtgagg atatgcagtt ttctttgtat atcattcttc ttcttctttg tagcttggag 1320

tcaaaatcgg ttccttcatg tacatacatc aaggatatgt ccttctgaat ttttatatct 1380

tgcaataaaa atgcttgtcc caaaaaaaaa aaaaaaaaaa aaa 1423

<210> 60
<211> 1841
<212> DNA
<213> Oryza sativa

<400> 60
ctcaccgaca ccatccgtag gtcttccagg agctccttcc tgccacgtca tcaatggcga 60

tgatgggtgg ctgacagtca aacgctcccc acgcctcctc cccttccccc ctctctccct 120

ccatggcttc cctcgcctcc cctcctctcc cctgccgcgc cgccgccacc gccagccgca 180

gcgggcgtcc tgctccgcgc ctcctcggcc ctccgccgcc gcccgcttcc cctctcctct 240

cctccgcttc ggcgcgcttc ccgcgtgccc cctgcaacgc cgcacgctgg agccggcgcg 300

acgccgtgcg ggtttgctct caagctggtg cagctggacc agccccatta tcgaagacat 360

tgtcagacct caaggattcc tgctggagat ttttacggcc acatacaatt cgaggaactg 420

ccttgggatc catagcatta gttgctagag ctttgataga gaaccccccaa ctgataaatt 480

ggtggttggt attcaaagcg ttctatgggc tcgtggcgtt aatctgtggc aatggttaca 540

tcgttgggat caatcagatc tatgacatta gaatcgataa ggtaaacaag ccatatttac 600

caattgctgc cggtgatctc tcagttcaga cagcatggtt attggtggta ttatttgcag 660

ctgcgggatt ttcaattgtt gtgacaaact ttatactttt cattacctct ctatactgcc 720

ttgctctatt tcttggcacc atatactctg ttcctccatt cagacttaag agatatcgtg 780

cgcctgcatg ccttatcatt gcaacggtcc gcggttttct ccgcaacttg ggtgtgtact 840

atgctactag agcagcactg ggtcttacat tccaatggag ctcgcctgtt gctttcatta 900

catgcttcgt gactttattt gctttggtca ttgctataac caaagatctc ccagatgttg 960

aaggggatcg gaagtatcaa atatcaactt tggcgacaaa gctcggtgtc agaaacattg 1020

catttcttgg ctctggttta ttgatagcaa attatgttgc tgctattgct gtagcttttc 1080

tcatgcctca ggctttcagg cgcactgtaa tggtgcctgt gcatgctgcc cttgccgttg 1140

gtataatttt ccagacatgg gttctggagc aagcaaaata tactaaggat gctatttcac 1200

agtactaccg gttcatttgg aatctcttct atgctgaata catcttcttc ccgttgatat 1260

agagaccaag caatctgata tggtctgcat gttgagtgcg gcaaaaacta gaagcccata 1320

tgaacagtgg gagtaaggga acgaacatgc catccatggg aagactctga taactctctc 1380

tcgcccgggc tgtaaagggt aagcactgtt gtgcatatat atgaaaggaa ggtgataaag 1440

cagggatgct aaattgctac tgggatcctc aaaggcttat agtggtcatc agtggaatgt 1500

gccttaataa tttggttacc tagcagagca agtttttgca ggttattagg taatatcttt 1560

gagggaatga acttagattt cattgtttta aggtctggtc acacaacggg tagtagtgct 1620

ggagcggcaa aagacgacct tgttttacac taccaaggga ggttaactct agttttcatg 1680

tgaccactta ccttgagagt tgagaccatg gaatcacttg tcgactcctc ggcttgtata 1740

tttctagtgt cagcatttgc attctcctcc acacttgtac ttgaagagtt gaagacaact 1800

tttttgtttg tgtatttctg gagtgtcagc atttgcattc c 1841

<210> 61
<211> 970
<212> PRT
<213> Arabidopsis thaliana

<400> 61

Met Asp Pro Pro Val Ser Asp Leu Glu Ser Ile Glu Asp Gln Lys Glu
1 5 10 15

Gly Gly Pro Ser Phe His Cys Asp Leu Tyr Asp Thr Gln Val Val His
20 25 30

Lys Ile Ala Gln Val Phe Leu Pro Gly Leu Ala Thr Ala Cys Val Asp
35 40 45

Asn Thr Thr Gly Asp Ile Phe Arg Ser Pro Gly Ser Val Ala Ala Asp
50 55 60

Ile Arg Lys Glu Met Ile Glu Tyr Leu Thr Arg Arg Ser Glu Thr Phe
65 70 75 80

Val Ala Glu His Ile Val Leu Gln Gly Gly Ser Glu Ile Glu Ala Ser
85 90 95

His Asp Pro Phe Asp Ile Ile Ser Asp Phe Ile Asp Asp Phe Ala Thr
100 105 110

Ser Lys Arg Asn Leu Phe Ser Arg Val Ser Gly Trp Met Leu Ser Glu
115 120 125

Arg Arg Glu Asp Asn Ile Asp Asp Phe Ala Gln Glu Met Glu Ile Ser
130 135 140

Gly Phe Trp Leu Thr Asp His Arg Glu Gly Ile Ala Gln Thr Leu Leu
145 150 155 160

Lys Asn Val Asp Phe Lys Ser Ser Ala His Cys Glu Met Lys Phe Gln
165 170 175

Thr Glu Gly Glu Leu Ala Glu His Ala Met Asn Cys Gly Tyr Arg Thr
180 185 190

Met Asn Cys Glu Asn Glu Gly Cys Thr Ala Val Phe Cys Ala Asn Gln
195 200 205

Met Glu Asn His Asp Ser Val Cys Pro Phe Lys Ile Ile Pro Cys Glu
210 215 220

Gln Asn Cys Ser Glu Ser Ile Met Arg Arg Asp Met Asp Arg His Cys
225 230 235 240

Ile Thr Val Cys Pro Met Lys Leu Val Asn Cys Pro Phe His Ser Val
245 250 255

Gly Cys Leu Ser Asp Val His Gln Cys Glu Val Gln Gln His His Leu
260 265 270

Asp Asn Val Ser Ser His Leu Met Tyr Ile Leu Arg Ser Ile Tyr Lys
275 280 285

Glu Ala Ser Leu Asp Asp Leu Lys Pro Arg Ala Glu Gln Ile Gln Gln
290 295 300

Leu Ser Thr Arg Leu Ser Glu Ala Arg Asn Ala Arg Ser Leu Thr Asn
305 310 315 320

Leu Val Lys Glu Ile Asp Gly Lys Leu Gly Pro Leu Glu Ile Lys Pro
325 330 335

Lys Ile Val Thr Asp Ser Glu Ser Asp Lys Pro Glu Asn Thr Glu Lys
340 345 350

Lys Ala Leu Glu Glu Ala Glu Ile Lys Glu Lys Pro Glu Thr Ser Asn
355 360 365

Leu Lys Ala Val Thr Leu Glu Gln Thr Ala Arg Glu Ala Pro Glu Asp
370 375 380

Lys Leu Val Ser Lys Glu Val Asp Ala Ala Met Val Lys Glu Ala Ala
385 390 395 400

Lys Lys Val Ser Glu Ala Glu Ile Ala Asp Asn Val Asn Glu Glu Gly
405 410 415

Glu Leu Lys Ala Gln Lys Leu Leu Glu Ile Gly Glu Phe Ile Lys Glu
420 425 430

Gly Asp Asn Asn Ser Ala Asp Asp Leu Ser Glu Arg Thr Glu Thr Lys
435 440 445

Ala Pro Glu Val Val Val Met Asp Glu Ala Arg Glu Glu Glu Asp Ser
450 455 460

Val Glu Thr Lys Asp Thr Arg Thr Tyr Glu Thr Ile Arg Gly Leu Glu
465 470 475 480

Ile Glu Ala Asn Glu Met Ile Asp Glu Glu Thr Lys Lys Ser Thr Glu
485 490 495

Thr Lys Thr Glu Ala Pro Ser Arg Ile Val Met Asp Lys Glu Gly Asp
500 505 510

Glu Glu Thr Lys Lys Ser Thr Glu Thr Glu Thr Glu Ala Pro Ser Arg
515 520 525

Ile Val Met Glu Thr Glu Lys Asp Glu Glu Thr Met Asn Ser Arg Ala
530 535 540

Arg Ala Ser Asp Glu Ala Glu Ala Leu Ser Lys Ser Ser Gln Val Ala
545 550 555 560

Ser Met Glu Leu Ser Ile Ser Gln Ser Pro Arg Val Arg Phe Ser Ser
565 570 575

Leu Ala Pro Arg Phe Leu Ala Ala Ser His His His Arg Pro Ser Val
580 585 590

His Leu Ala Gly Lys Phe Ile Ser Leu Pro Arg Asp Val Arg Phe Thr
595 600 605

Ser Leu Ser Thr Ser Arg Met Arg Ile Leu Ala Val Ala Leu Thr Phe
610 615 620

Lys Ser Arg Cys Val Tyr Val Asn Tyr Glu Ile Pro Lys Asp Gln Ile
625 630 635 640

Leu Val Gly Ala Ala Glu Ser Asp Asp Pro Val Leu Asp Arg Ile Ala
645 650 655

Arg Phe Gln Asn Ala Cys Trp Arg Phe Leu Arg Pro His Thr Ile Arg
660 665 670

Gly Thr Ala Leu Gly Ser Thr Ala Leu Val Thr Arg Ala Leu Ile Glu
675 680 685

Asn Thr His Leu Ile Lys Trp Ser Leu Val Leu Lys Ala Leu Ser Gly
690 695 700

Leu Leu Ala Leu Ile Cys Gly Asn Gly Tyr Ile Val Gly Ile Asn Gln
705 710 715 720

Ile Tyr Asp Ile Gly Ile Asp Lys Val Asn Lys Pro Tyr Leu Pro Ile
725 730 735

Ala Ala Gly Asp Leu Ser Val Gln Ser Ala Trp Leu Leu Val Ile Phe
740 745 750

Phe Ala Ile Ala Gly Leu Leu Val Val Gly Phe Asn Phe Gly Pro Phe
755 760 765

Ile Thr Ser Leu Tyr Ser Leu Gly Leu Phe Leu Gly Thr Ile Tyr Ser
770 775 780

Val Pro Pro Leu Arg Met Lys Arg Phe Pro Val Ala Ala Phe Leu Ile
785 790 795 800

Ile Ala Thr Val Arg Gly Phe Leu Leu Asn Phe Gly Val Tyr His Ala
805 810 815

Thr Arg Ala Ala Leu Gly Leu Pro Phe Gln Trp Ser Ala Pro Val Ala
820 825 830

Phe Ile Thr Ser Phe Val Thr Leu Phe Ala Leu Val Ile Ala Ile Thr

835 840 845

Lys Asp Leu Pro Asp Val Glu Gly Asp Arg Lys Phe Gln Ile Ser Thr
850 855 860

Leu Ala Thr Lys Leu Gly Val Arg Asn Ile Ala Phe Leu Gly Ser Gly
865 870 875 880

Leu Leu Leu Val Asn Tyr Val Ser Ala Ile Ser Leu Ala Phe Tyr Met
885 890 895

Pro Gln Tyr Ala Ala Leu Lys Arg Pro Thr Leu Leu Ser Phe Asn Asn
900 905 910

Glu Gln Val Phe Arg Gly Ser Leu Met Ile Pro Ala His Val Ile Leu
915 920 925

Ala Ser Gly Leu Ile Phe Gln Thr Trp Val Leu Glu Lys Ala Asn Tyr
930 935 940

Thr Lys Glu Ala Ile Ser Gly Tyr Tyr Arg Phe Ile Trp Asn Leu Phe
945 950 955 960

Tyr Ala Glu Tyr Leu Leu Phe Pro Phe Leu
965 970

<210> 62
<211> 441
<212> PRT
<213> Arabidopsis thaliana

<400> 62

Met Glu Leu Ser Ile Ser Gln Ser Pro Arg Val Arg Phe Ser Ser Leu
1 5 10 15

Ala Pro Arg Phe Leu Ala Ala Ser His His His Arg Pro Ser Val His
20 25 30

Leu Ala Gly Lys Phe Ile Ser Leu Pro Arg Asp Val Arg Phe Thr Ser
35 40 45

Leu Ser Thr Ser Arg Met Arg Ser Lys Phe Val Ser Thr Asn Tyr Arg
50 55 60

Lys Ile Ser Ile Arg Ser Val Cys Ala Phe Cys Asn Gly Thr His Lys
65 70 75 80

Ser Arg Tyr Tyr Gln Ala Cys Ser Gln Val Gly Ala Ala Glu Ser Asp
85 90 95

Asp Pro Val Leu Asp Arg Ile Ala Arg Phe Gln Asn Ala Cys Trp Arg
100 105 110

Phe Leu Arg Pro His Thr Ile Arg Gly Thr Ala Leu Gly Ser Thr Ala
115 120 125

Leu Val Thr Arg Ala Leu Ile Glu Asn Thr His Leu Ile Lys Trp Ser
130 135 140

Leu Val Leu Lys Ala Leu Ser Gly Leu Leu Ala Leu Ile Cys Gly Asn
145 150 155 160

Gly Tyr Ile Val Gly Ile Asn Gln Ile Tyr Asp Ile Gly Ile Asp Lys
165 170 175

Val Asn Lys Pro Tyr Leu Pro Ile Ala Ala Gly Asp Leu Ser Val Gln
180 185 190

Ser Ala Trp Leu Leu Val Ile Phe Phe Ala Ile Ala Gly Leu Leu Val
195 200 205

Val Gly Phe Asn Phe Gly Pro Phe Ile Thr Ser Leu Tyr Ser Leu Gly
210 215 220

Leu Phe Leu Gly Thr Ile Tyr Ser Val Pro Pro Leu Arg Met Lys Arg
225 230 235 240

Phe Pro Val Ala Ala Phe Leu Ile Ile Ala Thr Val Arg Gly Phe Leu
245 250 255

Leu Asn Phe Gly Val Tyr His Ala Thr Arg Ala Ala Leu Gly Leu Pro
260 265 270

Phe Gln Trp Ser Ala Pro Val Ala Phe Ile Thr Ser Phe Val Thr Leu
275 280 285

Phe Ala Leu Val Ile Ala Ile Thr Lys Asp Leu Pro Asp Val Glu Gly
290 295 300

Asp Arg Lys Phe Gln Ile Ser Thr Leu Ala Thr Lys Leu Gly Val Arg
305 310 315 320

Asn Ile Ala Phe Leu Gly Ser Gly Leu Leu Leu Val Asn Tyr Val Ser
325 330 335

Ala Ile Ser Leu Ala Phe Tyr Met Pro Gln Tyr Ala Ala Leu Lys Arg
340 345 350

Pro Thr Leu Leu Ser Phe Asn Asn Glu Gln Val Phe Arg Gly Ser Leu
355 360 365

Met Ile Pro Ala His Val Ile Leu Ala Ser Gly Leu Ile Phe Gln Thr
370 375 380

Trp Val Leu Glu Lys Ala Asn Tyr Thr Lys Ser Ile Cys Tyr Ser Pro
385 390 395 400

Ser Ser Ser Phe Gln Phe His Gly Glu Asp Met Gln Phe Ser Leu Tyr
405 410 415

Ile Ile Leu Leu Leu Leu Cys Ser Leu Glu Ser Lys Ser Val Pro Ser
420 425 430

Cys Thr Tyr Ile Lys Asp Met Ser Phe
435 440

<210> 63
<211> 155
<212> DNA
<213> Arabidopsis thaliana

<400> 63
gtgtggctag tggcatcaat ggagctctcg atctcacaat caccgcgtgt tcggttctcg 60

tctctggcgc ctcgtttctt agcagcttct catcatcatc gtccttctgt gcatttagct 120

gggaagttta taagcctccc tcgagatgtt cgctt 155

<210> 64
<211> 626
<212> DNA
<213> Medicago truncatula

<400> 64
gctccttatg gagctctcta cttcttcatc atcttctctt cattcacatt ccataattcc 60

cacatggaat tccaaaaact actactcttt caaaccaccc atttcagcta agtccacaac 120

cccaaaatct tccaaacggt ttggttcaat tgggctgcac catcatcatc acacaagttt 180

ctctgctcat gtttcaaaac cgaagagaca gtgtaaaccc atttccatca gggcctgcag 240

tgaagttgga gctgctggat ctgatcgtcc atttgctgac aaagttttag attttaaaga 300

tgcattctgg agatttttaa ggccacatac tatccgtggg acagcattag gctcttttgc 360

tttggtgtca agagcgttga ttgagaactc aaatctgata aagtggtctc ttttgttgaa 420

agcactctct ggacttttg ctctgatttg tgggaatggt tatatagttg gcatcaatca 480

aatatatgat atcggcattg acaaggtaaa caaaccttat ttacctatag ctgcaggaga 540

tctttctgtc caatctgcat ggtacttggt tatattcttt gcagcagctg gcctttgac 600

tgtaggattg aactttggat ctttta 626

<210> 65
<211> 651
<212> DNA
<213> Medicago truncatula

<400> 65
cttcagctcc ttatggagct ctctacttct tcatcatctt ctcttcattc acattccata 60

attcccacat ggaattccaa aaactactac tctttcaaac cacccatttc agctaagtcc 120

acaaccccaa aatcttccaa acggtttggt tcaattgggc tgcaccatca tcatcacaca 180

agtttctctg ctcatgtttc aaaaccgaag agacagtgta aacccatttc catcagggcc 240

tgcagtgaag ttggagctgc tggatctgat cgtccatttg ctgacaaagt tttagatttt 300

aaagatgcat tctggagatt tttaaggcca catactatcc gtgggacagc attaggctct 360

tttgctttgg tgtcaagagc gttgattgag aactcaaatc tgataaagtg gtctcttttg 420

ttgaaagcac tctctggact ttttgctctg atttgtggga atggttatat agttggcatc 480

aatcaaatat atgatatcgg cattgacaag gtaaacaaac cttatttacc tatagctgca 540

ggagatcttt ctgtccaatc tgcatggtac ttggttatat tctttgcagc agctggcctt 600

ttgactgtag gattgaactt tggatcttta ttttttctct ttactccttc g 651

<210> 66
<211> 651
<212> DNA
<213> Medicago truncatula

<400> 66
cttcagctcc ttatggagct ctctacttct tcatcatctt ctcttcattc acattccata 60

attcccacat ggaattccaa aaactactac tctttcaaac cacccatttc agctaagtcc 120

acaaccccaa aatcttccaa acggtttggt tcaattgggc tgcaccatca tcatcacaca 180

agtttctctg ctcatgtttc aaaaccgaag agacagtgta aacccatttc catcagggcc 240

tgcagtgaag ttggagctgc tggatctgat cgtccatttg ctgacaaagt tttagatttt 300

aaagatgcat tctggagatt tttaaggcca catactatcc gtgggacagc attaggctct 360

tttgctttgg tgtcaagagc gttgattgag aactcaaatc tgataaagtg gtctcttttg 420

ttgaaagcac tctctggact ttttgctctg atttgtggga atggttatat agttggcatc 480

aatcaaatat atgatatcgg cattgacaag gtaaacaaac cttatttacc tatagctgca 540

ggagatcttt ctgtccaatc tgcatggtac ttggttatat tctttgcagc agctggcctt 600

ttgactgtag gattgaactt tggatcttta ttttttctct ttactccttc g 651

<210> 67
<211> 614
<212> DNA
<213> Medicago truncatula

<400> 67
gttgagtgtt gcttcagctc cttatggagc tctctacttc ttcatcatct tctcttcatt 60

cacattccat aattcccaca tggaattcca aaaactacta ctctttcaaa ccacccattt 120

cagctaagtc cacaacccca aaatcttcca aacggtttgg ttcaattggg ctgcaccatc 180

atcatcacac aagtttctct gctcatgttt caaaaccgaa gagacagtgt aaacccattt 240

ccatcagggc ctgcagtgaa gttggagctg ctggatctga tcgtccattt gctgacaaag 300

ttttagattt taaagatgca ttctggagat ttttaaggcc acatactatc cgtgggacag 360

cattaggctc ttttgctttg gtgtcaagag cgttgattga gaactcaaat ctgataaagt 420

ggtctctttt gttgaaagca ctctctggac tttttgctct gatttgtggg aatggttata 480

tagttggcat caatcaaata tatgatatcg gcattgacaa ggtaaacaaa ccttatttac 540

ctatagctgc aggagatctt tctgtaccaa tctgcatggt acttggttat attctttgca 600

acagctggcc tttt 614

<210> 68
<211> 560
<212> DNA
<213> Solanum tuberosum

<400> 68
gacccttttc ccatatattt atccacttac caccttatcc tcttgaggtt gaacaaattc 60

attcttcctt tggtatggag atacagggtg ttacaataga caaaattgac tcacacaaga 120

aacattccat ctcctaagga tagattgatg aatttcctaa gaatatacaa caacaggata 180

aataaattct ccatcccccc catctctata gacaattttc ctttgcaaac cagaaagaaa 240

acttgcattt cttcgaggtc catcgaacaa aaacctttat caagtagatt ttagatgaaa 300

ggaaagatga tgtactccgc atagaaaaga ttccatataa atctatagta tgctgagatt 360

gcttccttgg tgtaatttgc tctttccaac aaccatgcct ggaaaactaa acacgatgct 420

aagatgacat gtacaggtat catcaagcta cacctgaatg cttggggcat gtatattgct 480

gcaactacag caccaatgta atttgttaat aatagaccag aaccaaggaa tgctatgttt 540

ctcacaccaa gctttgttgc 560

<210> 69
<211> 371
<212> DNA
<213> Arabidopsis thaliana

<400> 69
caattggtac aagcattttt attgcaagat ataaaaattc agaaggacat atccttgatg 60

tatgtacatg aaggaaccga ttttgactcc aagctacaaa gaagaagaag aatgatatac 120

aaagaaaact gcatatcctc accatgaaat tgaaagctag aggaagggga ataacagata 180

ctctgcgtag aagagattcc atataaaccg ataatatcct gagatagctt ccttggtgta 240

gtttgctttt tctagtaccc atgtctggaa aattaagcct gaagccaaga tcacatgtgc 300

aggaatcatc aagctacctc taaaaacctg aggcatgtag aaagctagtg atatggctga 360

aacataattt a 371

<210> 70
<211> 357
<212> DNA
<213> Arabidopsis thaliana

<400> 70
ggtacaagca tttttattgc aagatataaa aattcagaag gacatatcct tgatgtatgt 60

acatgaagga accgattttg actccaagct acaaagaaga agaagaatga tatacaaaga 120

aaactgcata tcctcaccat gaaattgaaa gctagaggaa ggggaataac agatactctg 180

cgtagaagag attccatata aaccgataat atcctgagat agcttccttg gtgtagtttg 240

ctttttctag tacccatgtc tggaaaatta agcctgaagc caagatcaca ctggttgcag 300

gaactcatca agctaccctc taaaaaaacc gtgaggcaac tgctaggaaa acaggcc 357

<210> 71
<211> 560
<212> DNA
<213> Medicago truncatula

<220>
<221> misc_feature
<222> (537)..(537)
<223> unknown

<400> 71
tgccacaggc tttcaggaga tggttactga taccagctca tgcaatattt gcttcaagct 60

taatttacca ggtgcagata ttagaaaaag caaattatac aaaggaagca atatcaggat 120

tctatcgatt catatggaat ctgttctatg ccgagtatgc attatttcct ttcatctagc 180

aaactgtgct acatttttac ttggaaaaat tgcacacatg catccaaaaa tgcagcggtt 240

gcttgaccaa agccggtcaa taagacaaag ccgttcaata agaaaaatct tagttatatc 300

gagtatctat tcttaaagta ttaacaattt tttttaatgg tttgagtaaa tttttgtata 360

tagtatagtg cttccttta atgagatgta ttgccatgag aattgtatac aacggccaga 420

tttcatttgt gttggaacaa attccactgg tgaatgtgat aatatactca tgtgaactct 480

acccaaaaat aaaataaaat aaaaaaaaaa aaaaaaaaaa aaaaataaaa aaaaaanaaa 540

aataaaaaaa cgtcgagggg 560

<210> 72
<211> 713
<212> DNA
<213> Glycine max

<220>
<221> misc_feature
<222> (13)..(13)
<223> unknown

<220>
<221> misc_feature
<222> (679)..(679)
<223> unknown

<220>
<221> misc_feature
<222> (681)..(683)
<223> unknown

<220>
<221> misc_feature
<222> (686)..(693)
<223> unknown

<220>
<221> misc_feature
<222> (28)..(29)
<223> unknown

<220>
<221> misc_feature
<222> (634)..(635)
<223> unknown

<220>
<221> misc_feature
<222> (637)..(637)

<223> unknown

<220>
<221> misc_feature
<222> (643)..(643)
<223> unknown

<220>
<221> misc_feature
<222> (658)..(660)
<223> unknown

<220>
<221> misc_feature
<222> (664)..(664)
<223> unknown

<220>
<221> misc_feature
<222> (669)..(670)
<223> unknown

<220>
<221> misc_feature
<222> (672)..(674)
<223> unknown

<400> 72
aaatgcgatg ganagccgat gacgcctnna tatacaaaaa tactcattaa aaaaatagct 60

agtttactaa ttgatacttg agacaacaaa gaagatctct ctaactatgc acgtatgcac 120

cttttcccaa gaaaaatgta gcaaggcttg ctatatgaaa ggaaatattg catactcagc 180

atagaacaga ttccatatga atcgatagaa tcctgatatt gcatccttgg tataatttgc 240

ttgttctaat atccatgcct ggtaaatcaa gcttattgca aaaatcgtat gagccggtat 300

gagtaaccaa cgcctgaaag cctgaggcat ataaattgct gccaaaacag aaacaatata 360

gttcaccaac aaaattccag aaccaaggaa agcaatgttc cgaactccta attttgtggc 420

aaaggttgat atctgatact tgcgatctcc ttcaacatca ggaagatctt ttgttatagc 480

aattaccagt gcgaaaaatg ttacaaatgt tgtgataaaa accacaggag agctccattc 540

aaatgcaagc ccaagggcag ctctagtggc atagtacaca ccaaagttaa agagaaaacc 600

ccgaaccgtg gcaattataa gaaatgctgc aacnngnaag cgnttcattc taaatggnnn 660

aacngaatnn annntaccna nnnaannnnn nnntgtgtaa agagaaaaaa tga 713

<210> 73
<211> 56

<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 73
cgcgatttaa atggcgcgcc ctgcaggcgg ccgcctgcag ggcgcgccat ttaaat 56

<210> 74
<211> 32
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 74
tcgaggatcc gcggccgcaa gcttcctgca gg 32

<210> 75
<211> 32
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 75
tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 76
<211> 32
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 76
tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 77
<211> 32
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 77
tcgaggatcc gcggccgcaa gcttcctgca gg 32

<210> 78
<211> 36
<212> DNA

<213> Artificial

<220>
<223> Primer

<400> 78
tcgaggatcc gcggccgcaa gcttcctgca ggagct 36

<210> 79
<211> 28
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 79
cctgcaggaa gcttgcggcc gcggatcc 28

<210> 80
<211> 36
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 80
tcgacctgca ggaagcttgc ggccgcggat ccagct 36

<210> 81
<211> 28
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 81
ggatccgcgg ccgcaagctt cctgcagg 28

<210> 82
<211> 39
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 82
gatcacctgc aggaagcttg cggccgcgga tccaatgca 39

<210> 83
<211> 31
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 83
ttggatccgc ggccgcaagc ttcctgcagg t 31

<210> 84
<211> 31
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 84
cacatatggc atgttctcag gttggtgctg c 31

<210> 85
<211> 29
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 85
gcgtcgacct agaggaaggg gaataacag 29

<210> 86
<211> 31
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 86
caaccatggc atgttctcag gttggtgctg c 31

<210> 87
<211> 29
<212> DNA
<213> Artificial

<220>
<223> Primer

<400> 87
gcgtcgacct agaggaaggg gaataacag 29

<210> 88
<211> 1290
<212> DNA
<213> Trichodesmium erythraeum

<400> 88
atgggaaaaa ttgctggttc tcaacaggga aaaattacaa cgaattggct acaaaaatat 60

gtgccatggc tttatagttt ttggaagttt gctcgcccac atacgattat tggtacatcg 120

ttaagtgtgc tggctttata tataattgcc atgggcgatc gctctaattt ttttgacaaa 180

tatttttttt tatacagctt aattctgtta ttgataactt ggattagttg tttatgtgga 240

atgggaaaaa ttgctggttc tcaacaggga aaaattacaa cgaattggct acaaaaatat 300

gtgccatggc tttatagttt ttggaagttt gctcgcccac atacgattat tggtacatcg 360

ttaagtgtgc tggctttata tataattgcc atgggcgatc gctctaattt ttttgacaaa 420

tatttttttt tatacagctt aattctgtta ttgataactt ggattagttg tttatgtgga 480

aatatttata tagtaggatt aaatcaatta gaggatatag aaatagatag gattaataag 540

cctcatctcc ctatagctgc tggtgagttt tctcgttttt ctggtcaaat aattgtggta 600

ataacgggta ttttggcttt gagttttgcc gggttggggg gacctttttt gttgggtaca 660

gtggggataa gtttggcaat tggtacggct tattctttac ctcctattcg attaaaaaga 720

tttcctgttt tggccgcatt atgtattttt actgtgcggg gagttattgt taatttgggt 780

atatttttaa gctttgtttg ggggtttgaa aaggttgagg aggtttcagg aggtttaatt 840

aaatggatgg gtgagttggg tgaggttgtt ctacttcaaa aaagcttgat ggttccagaa 900

attcctctga cggtatgggc tttaactttg tttgtgatag tatttacttt tgctattgct 960

atttttaagg atattccaga tattgagggt gaccgtcaat ataatattaa tacgtttacg 1020

attaagttgg gagcatttgc tgtttttaat ttggcaaggt gggtattgac tttttgctat 1080

ctgggtatgg tgatggtggg tgtagtttgg ttggcgagtg ttaatttatt ttttttggtg 1140

attagtcatt tattggcttt gggtataatg tggtggttta gtcaaagggt agatctgcat 1200

gataaaaagg cgatcgctga tttttatcaa tttatttgga aattattttt cctggaatat 1260

ctaatttttc ctatggcttg tttttttaa 1290

<210> 89
<211> 903
<212> DNA
<213> Chloroflexus aurantiacus

<400> 89
atgcgcaaac agctacgcct gctcattgaa tttgcccgtc cccacaccgt cattgctacc 60

agcgtccagg ttctgaccat gctgatcatc gtgatcggct ggcacccacc aacgctcgaa 120

ctggtgggac tggtcggggt gacgctcgtt gtctgtctgg cgctcaatct ctacgtagtc 180

ggcgtgaatc aactgaccga tgtggcgatt gatcggatca acaagccatg gctaccggtt 240

gctgccggtc agctttcatc ggatgctgcg caacgtattg ttatcagtgc cctgtttatt 300

gccctgaccg gtgcggctat gctcggccca ccgctctggt ggacggtgag tatcatcgcg 360

ctgatcggtt cactctactc gctccccccg ctgcgcttga agcgtcatcc cctcgctgcg 420

gccctcagta ttgccggtgc ccgcggggtg attgccaatc tcggcctggc cttccactat 480

cagtactggt tagatagcga attgccgatc acgaccctga tcctggtggc aaccttcttt 540

ttcggtttcg ctatggtgat cgcgctctat aaagacttgc ccgatgatcg cggtgatcgg 600

ttgtatcaga tcgagaccct gaccacgcgc ctcggcccgc agcgagtgct gcacctgggc 660

agaatcttgc tcaccgcctg ttatctgctt ccgattgccg tcggtctctg gtcgctgccg 720

acttttgccg ccgcgttcct ggccctcagc catgtggtcg ttatcagtgt tttctggctg 780

gtcagtatgc gcgttgatct gcaacgccgg caatcgattg ccagttttta tatgtttctg 840

tgggggattt tttataccga atttgccctg cttagcattt atcgtctgac gtataccctc 900

tga 903

<210> 90
<211> 389
<212> PRT
<213> Glycine max

<400> 90

Met Glu Leu Ser Leu Ser Pro Thr Ser His Arg Val Pro Ser Thr Ile
1 5 10 15

Pro Thr Leu Asn Ser Ala Lys Leu Ser Ser Thr Lys Ala Thr Lys Ser
20 25 30

Gln Gln Pro Leu Phe Leu Gly Phe Ser Lys His Phe Asn Ser Ile Gly
35 40 45

Leu His His His Ser Tyr Arg Cys Cys Ser Asn Ala Val Pro Glu Arg
50 55 60

Pro Gln Arg Pro Ser Ser Ile Arg Ala Cys Thr Gly Val Gly Ala Ser
65 70 75 80

Gly Ser Asp Arg Pro Leu Ala Glu Arg Leu Leu Asp Leu Lys Asp Ala
85 90 95

Cys Trp Arg Phe Leu Arg Pro His Thr Ile Arg Gly Thr Ala Leu Gly
100 105 110

Ser Phe Ala Leu Val Ala Arg Ala Leu Ile Glu Asn Thr Asn Leu Ile
115 120 125

Lys Trp Ser Leu Phe Phe Lys Ala Phe Cys Gly Leu Phe Ala Leu Ile
130 135 140

Cys Gly Asn Gly Tyr Ile Val Gly Ile Asn Gln Ile Tyr Asp Ile Ser
145 150 155 160

Ile Asp Lys Val Asn Lys Pro Tyr Leu Pro Ile Ala Ala Gly Asp Leu
165 170 175

Ser Val Gln Ser Ala Trp Phe Leu Val Ile Phe Phe Ala Ala Ala Gly
180 185 190

Leu Ser Ile Ala Gly Leu Asn Phe Gly Pro Phe Ile Phe Ser Leu Tyr
195 200 205

Thr Leu Gly Leu Phe Leu Gly Thr Ile Tyr Ser Val Pro Pro Leu Arg
210 215 220

Met Lys Arg Phe Pro Val Ala Ala Phe Leu Ile Ile Ala Thr Val Arg
225 230 235 240

Gly Phe Leu Leu Asn Phe Gly Val Tyr Tyr Ala Thr Arg Ala Ser Leu
245 250 255

Gly Leu Ala Phe Glu Trp Ser Ser Pro Val Val Phe Ile Thr Thr Phe
260 265 270

Val Thr Phe Phe Ala Leu Val Ile Ala Ile Thr Lys Asp Leu Pro Asp
275 280 285

Val Glu Gly Asp Arg Lys Tyr Gln Ile Ser Thr Phe Ala Thr Lys Leu
290 295 300

Gly Val Arg Asn Ile Ala Phe Leu Gly Ser Gly Ile Leu Leu Val Asn
305 310 315 320

Tyr Ile Val Ser Val Leu Ala Ala Ile Tyr Met Pro Gln Ala Phe Arg
325 330 335

Arg Trp Leu Leu Ile Pro Ala His Thr Ile Phe Ala Ile Ser Leu Ile
340 345 350

Tyr Gln Ala Arg Ile Leu Glu Gln Ala Asn Tyr Thr Lys Asp Ala Ile
355 360 365

Ser Gly Phe Tyr Arg Phe Ile Trp Asn Leu Phe Tyr Ala Glu Tyr Ala
370 375 380

Ile Phe Pro Phe Ile
385

<210> 91
<211> 1427
<212> DNA
<213> Glycine max

<400> 91
ttgcaatgta acattatgaa atatgttaat ggcattacgt caaagtaaaa ggaaatagta 60

tcacatttat atacaacaat actcatttta aaaaaataat gatagctagt ttaccaattg 120

acacttgata aacaaagatc tctctaacta tgcacgtatg caccttttcc caagaaaaaa 180

gtagcaaggt ttgctatatg aaaggaaata ttgcatactc agcatagaac agattccata 240

tgaatcgata gaatcctgat attgcatcct tggtataatt tgcttgttct aatattcgtg 300

cctggtaaat caagcttatt gcaaaaattg tatgagctgg tatgagtaac caacgcctga 360

aagcctgagg catataaatt gctgccaaaa cagaaacaat ataattcacc agcaaaattc 420

cagaaccaag gaaagcaatg ttccgaactc ctaattttgt agcaaaggtt gatatctgat 480

acttgcgatc accttcaaca tcaggaagat cttttgttat agcaattacc agtgcgaaaa 540

atgttacaaa tgttgtgata aaaaccacag gagagctcca ttcaaatgca agcccaaggg 600

aagctctagt ggcatagtac acaccaaagt taaggagaaa accacgtacc gtggcaatta 660

taagaaatgc tgcaacagga aagcgtttca tcctcaatgg aggaacagaa tagatggttc 720

ccaagaaaag gccaagtgtg taaagagaaa aaatgaaagg cccaaagttc aaccctgcaa 780

tcgacaggcc agctgctgca aaaaatataa ccaagaacca tgcagattgg acagaaagat 840

ctccagcagc tataggtaaa taaggtttgt ttaccttgtc aatgctaatg tcatagattt 900

gattgatgcc aactatataa ccattcccac aaatcagggc aaaaagacca cagaaagctt 960

tgaaaaaaag agaccacttt atcaaattcg tgttctcaat caatgctctt gccaccaaag 1020

caaatgaacc tagtgctgta ccacgtatag tatgtggcct taaaaatctc caacaagcat 1080

ctttcaaatc taaaagtctt tcagctaatg gacgatcaga tccagaagct ccaactccag 1140

tgcaagccct tatggaactg ggtctttggg gtctctcggg aacagcattt gagcagcatc 1200

tgtaactgtg atggtgcaac ccaattgagt tgaagtgttt ggaaaatcct aagaacaaag 1260

gttgttgtga cttggtggcc ttagtggatg atagtttagc ggaattcaaa gtgggaattg 1320

tggaaggaac acgatgtgaa gttggagaga gtgagagctc cataaggagc tgagcacagc 1380

aaacgagaaa acactccaaa tttcagacgc aacgcaaggc aaaaacc 1427

<210> 92
<211> 12
<212> PRT
<213> Artificial

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> x = w, i, or y

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> x = r, e, or k

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = l, a, or s

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = l, a, or s

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> x = i or v

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = i or r

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = g or a

<400> 92

Xaa Xaa Phe Xaa Arg Pro His Thr Xaa Xaa Xaa Thr
1 5 10

<210> 93
<211> 26

<212> PRT
<213> Artificial

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> x = v, i, l, or g

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> x = i, l, or v

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = i or l

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = s, f, y, e, w, or t

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> x = v or i

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> x = r, s, g, e, d, or a

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> x = k or r

<220>
<221> MISC_FEATURE
<222> (18)..(18)
<223> x = v or i

<220>
<221> MISC_FEATURE
<222> (22)..(22)
<223> x = y, d, t, w, n, or h

<220>

<221> MISC_FEATURE
<222> (25)..(25)
<223> x = i, v, or l

<400> 93

Asn Xaa Tyr Ile Val Gly Xaa Asn Gln Xaa Xaa Asp Xaa Xaa Ile Asp
1 5 10 15

Xaa Xaa Asn Lys Pro Xaa Leu Pro Xaa Ala
20 25

<210> 94
<211> 14
<212> PRT
<213> Artificial

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = l or i

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = f, t, or y

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> x = l, i, or v

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> x = i, v, m, or d

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> x = e, d, or r

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> x = r or k

<400> 94

Ile Ala Xaa Xaa Lys Asp Xaa Pro Asp Xaa Xaa Gly Asp Xaa
1 5 10

<210> 95
<211> 23
<212> PRT
<213> Artificial

<220>
<223> Conserved Motif

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> x = k or r

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> x = d, e, q, t, a, k, or s

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> x = a, e, s, or t

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> x = i or l

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> x = s, t, or a

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> x = q, g, d, or s

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> x = f, y, or c

<220>
<221> MISC_FEATURE
<222> (9)..(9)
<223> x = q, m, or r

<220>

<221> MISC_FEATURE
<222> (10)..(10)
<223> x = f or l

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> i, v, or l

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> x = n, k, or g

<220>
<221> MISC_FEATURE
<222> (14)..(14)
<223> x = l or i

<220>
<221> MISC_FEATURE
<222> (16)..(16)
<223> x = y or f

<220>
<221> MISC_FEATURE
<222> (17)..(17)
<223> x = a, l, i, or t

<220>
<221> MISC_FEATURE
<222> (19)..(19)
<223> x = y or f

<220>
<221> MISC_FEATURE
<222> (20)..(20)
<223> x = a, i, or l

<220>
<221> MISC_FEATURE
<222> (21)..(21)
<223> x = f, i, l, or m

<220>
<221> MISC_FEATURE
<222> (22)..(22)
<223> x = f, l, i, or y

<220>
<221> MISC_FEATURE

<222> (23)..(23)
<223> x = p or s

<400> 95

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Tyr Xaa Xaa Xaa Trp Xaa Xaa Phe Xaa
1 5 10 15

Xaa Glu Xaa Xaa Xaa Xaa Xaa
20

CLAIMS:

1. A transformed plant cell comprising an introduced nucleic acid molecule encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:57, wherein said plant cell comprises an increased total tocopherol level relative to a plant cell with a similar genetic background but lacking said introduced nucleic acid.

2. The transformed plant cell of claim 1, wherein the plant cell is from a plant selected from the group consisting of alfalfa, *Arabidopsis thaliana*, barley, *Brassica napus*, *Brassica campestris*, oilseed rape, broccoli, cabbage, citrus, canola, cotton, garlic, oat, Allium, flax, an ornamental plant, peanut, pepper, potato, rapeseed, rice, rye sorghum, strawberry, sugarcane, sugar beet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, soybean, chick peas, corn, Phaseolus, crambe, mustard, castor bean, sesame, cottonseed, linseed, safflower, and oil palm.

3. The transformed plant cell of claim 1, wherein said plant cell is from a plant selected from the group consisting of oilseed rape, soybean and canola.

4. The transformed plant cell of claim 1, wherein said transformed plant cell is from a plant which comprises tissue with an increased total tocopherol level relative to a plant with a similar genetic background but lacking said introduced nucleic acid molecule.

5. The transformed plant cell of claim 1, wherein said transformed plant cell is from a plant which produces a seed with an increased total tocopherol level relative to a plant with a similar genetic background but lacking said introduced nucleic acid molecule.

6. The transformed plant cell of claim 1, wherein said nucleic acid molecule is operably linked to a promoter.

7. The transformed plant cell of claim 6, wherein said promoter is a seed-specific promoter.

8. The transformed plant cell of claim 7, wherein said promoter is selected from the promoters consisting of: napin, 7S alpha, 7S alpha', USP 88, enhanced USP 88, Arcelin 5, and Oleosin.

9. A transformed plant cell comprising an introduced first nucleic acid molecule encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:57 or the complement thereof, and an introduced second nucleic acid molecule encoding an enzyme selected from the group consisting of bifunctional *prephenate*

dehydrogenase (tyrA), *tocopherol cyclase*, 1-deoxyxylulose-5-phosphate synthase (dxs), 1-deoxyxylulose-5-phosphate reductoisomerase (dxr), geranylgeranyl pyrophosphate synthase (GGPPS), p-hydroxyphenylpyruvate dioxygenase (HPPD), adenylate transporter (AANTI), isopentenyl diphosphate isomerase (IDI), geranylgeranyl diphosphate reductase (GGH), gamma methyl transferase (GMT), tocopherol methyl transferase 2 (tMT2), methyl transferase 1 (MT1), (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (gcpE) and complements thereof.

10. The transformed plant cell of claim 9, wherein said plant cell is from a plant selected from the group consisting of alfalfa, *Arabidopsis thaliana,* barley, *Brassica napus, Brassica campestris*, oilseed rape, broccoli, cabbage, citrus, canola, cotton, garlic, oat, Allium, flax, an ornamental plant, peanut, pepper, potato, rapeseed, rice, rye, sorghum, strawberry, sugarcane, sugar beet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, soybean, chick peas, corn, Phaseolus, crambe, mustard, castor bean, sesame, cottonseed, linseed, safflower, and oil palm.

11. The transformed plant cell of claim 9, wherein said plant cell is from a plant selected from the group consisting of canola, oilseed rape, and soybean.

12. The transformed plant cell of claim 9, wherein said plant cell is from a transformed plant which comprises tissue with increased α-tocopherol levels relative to a plant with a similar genetic background but lacking said introduced nucleic acid molecules.

13. The transformed plant cell of claim 9, wherein said transformed plant cell is from a transformed plant which produces a seed with an increased total tocopherol level relative to a plant with a similar genetic background but lacking said introduced nucleic acid molecules.

14. The transformed plant cell of claim 9, wherein at least one of said first and second nucleic acid molecules is operably linked to a promoter.

15. The transformed plant cell of claim 14, wherein said promoter is a seed specific promoter.

16. The transformed plant cell of claim 15, wherein said promoter is selected from the promoters consisting of: napin, 7S alpha, 7S alpha', USP 88, enhanced USP 88, Arcelin 5, and Oleosin.

17. A transformed plant cell comprising an introduced nucleic acid molecule comprising an introduced promoter region which functions in plant cells to cause the

production of an mRNA molecule, wherein said introduced promoter region is linked to a transcribed nucleic acid molecule having a transcribed strand and a non-transcribed strand, wherein said transcribed strand is complementary to a nucleic acid molecule encoding the polypeptide of SEQ ID NO:57 and wherein said transcribed nucleic acid molecule is linked to a 3' non-translated sequence that functions in the plant cells to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA sequence.

18. The transformed plant cell of claim 17, wherein the expression of homogentisate prenyl transferase is reduced relative to a plant with a similar genetic background but lacking said introduced nucleic acid molecule.

19. A method of producing a plant having a seed with an increased total tocopherol level comprising (A) transforming said plant with an introduced nucleic acid molecule encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:57 and (B) growing said transformed plant.

20. The method of producing a plant of claim 19, wherein said plant is selected from the group consisting of alfalfa, *Arabidopsis thaliana*, barley, *Brassica napus, Brassica campestris*, oilseed rape broccoli, cabbage, canola, citrus, cotton, garlic, oat, Allium, flax, an ornamental plant, peanut, pepper, potato, rapeseed, rice, rye, sorghum, strawberry, sugarcane, sugar beet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, soybean, corn, Phaseolus, crambe, mustard, castor bean, sesame, cottonseed, linseed, safflower, and oil palm.

21. The method of claim 19, wherein said plant is selected from the group consisting of canola, oilseed rape, and soybean.

22. The method of claim 19, wherein the plant is transformed with an introduced second nucleic acid molecule encoding an enzyme selected from the group consisting of bifunctional *prephenate dehydrogenase* (tyrA), *tocopherol cyclase*, 1-deoxyxylulose-5-phosphate synthase (dxs), 1-deoxyxylulose-t-phosphate reductoisomerase (dxr), geranylgeranyl pyrophosphate synthase (GGPPS), p-hydroxyphenylpyruvate dioxygenase (HPPD), adenylate transporter (AANTI), isopentenyl diphosphate isomerase (IDI), geranylgeranyl diphosphate reductase (GGH), gamma methyl transferase (GMT), tocopherol methyl transferase 2 (tMT2), methyl

transferase 1 (MT1), (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (gcpE) and complements thereof.

23. A seed cell derived from a seed of a transformed plant comprising an introduced nucleic acid molecule encoding a polypeptide comprising the amino acid of SEQ ID NO:57, wherein said seed cell comprises an increased total tocopherol level relative to a seed cell with a similar genetic background but lacking said introduced nucleic acid.

24. The seed cell of claim 23, wherein said seed has an increased total tocopherol level and relative to a seed from a plant having a similar genetic background but lacking said introduced nucleic acid molecule.

25. A seed cell derived from a seed of a transformed plant comprising an introduced first nucleic acid molecule encoding an introduced polypeptide comprising the amino acid sequence of SEQ ID NO:57 and an introduced second nucleic acid encoding an enzyme selected from the group consisting of bifunctional *prephenate dehydrogenase* (tyrA), *tocopherol cyclase*, 1-deoxyxylulose-5-phosphate synthase (dxs), 1-deoxyxylulose-t-phosphate reductoisomerase (dxr), geranylgeranyl pyrophosphate synthase (GGPPS), p-hydroxyphenylpyruvate dioxygenase (HPPD), adenylate transporter (AANTI), isopentenyl diphosphate isomerase (IDI), geranylgeranyl diphosphate reductase (GGH), gamma methyl transferase (GMT), tocopherol methyl transferase 2 (tMT2), methyl transferase 1 (MT1), (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (gcpE) and complements thereof.

26. The seed cell of claim 25, wherein said seed has at least one of an increased total tocopherol level and an increased tocotrienol level relative to a seed from a plant having a similar genetic background but lacking said introduced nucleic acid molecule.

27. The seed cell of claim 25, wherein the introduced second nucleic acid is *GMT*.

28. The seed cell of claim 27, wherein said seed has at least one of an increased total tocopherol level and an increased total tocotrienol level relative to a seed having a similar genetic background but lacking said introduced nucleic acid molecules, and at least about 90% of the total tocopherol is alpha tocopherol.

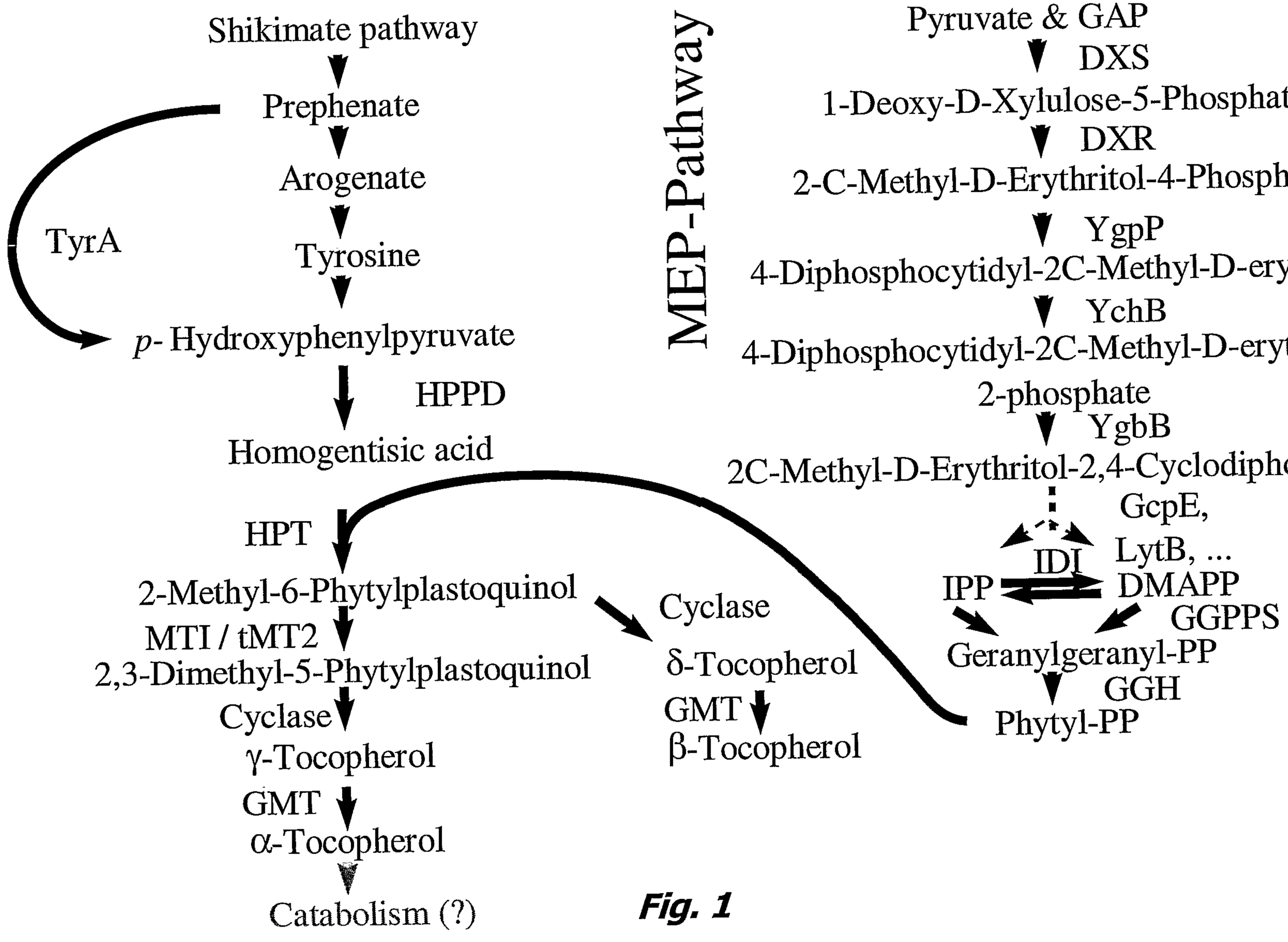

Shikimate pathway
Prephenate
Arogenate
TyrA
Tyrosine
p- Hydroxyphenylpyruvate
HPPD
Homogentisic acid
HPT
2-Methyl-6-Phytylplastoquinol
MTI / tMT2
2,3-Dimethyl-5-Phytylplastoquinol
Cyclase
γ-Tocopherol
GMT
α-Tocopherol
Catabolism (?)
Cyclase
δ-Tocopherol
GMT
β-Tocopherol
MEP-Pathway
Pyruvate & GAP
DXS
1-Deoxy-D-Xylulose-5-Phosphate
DXR
2-C-Methyl-D-Erythritol-4-Phosphate
YgpP
4-Diphosphocytidyl-2C-Methyl-D-erythritol
YchB
4-Diphosphocytidyl-2C-Methyl-D-erythritol
2-phosphate
YgbB
2C-Methyl-D-Erythritol-2,4-Cyclodiphosphate
GcpE,
LytB, ...
IDI
IPP
DMAPP
GGPPS
Geranylgeranyl-PP
GGH
Phytyl-PP

Fig. 1

```
                                                
```

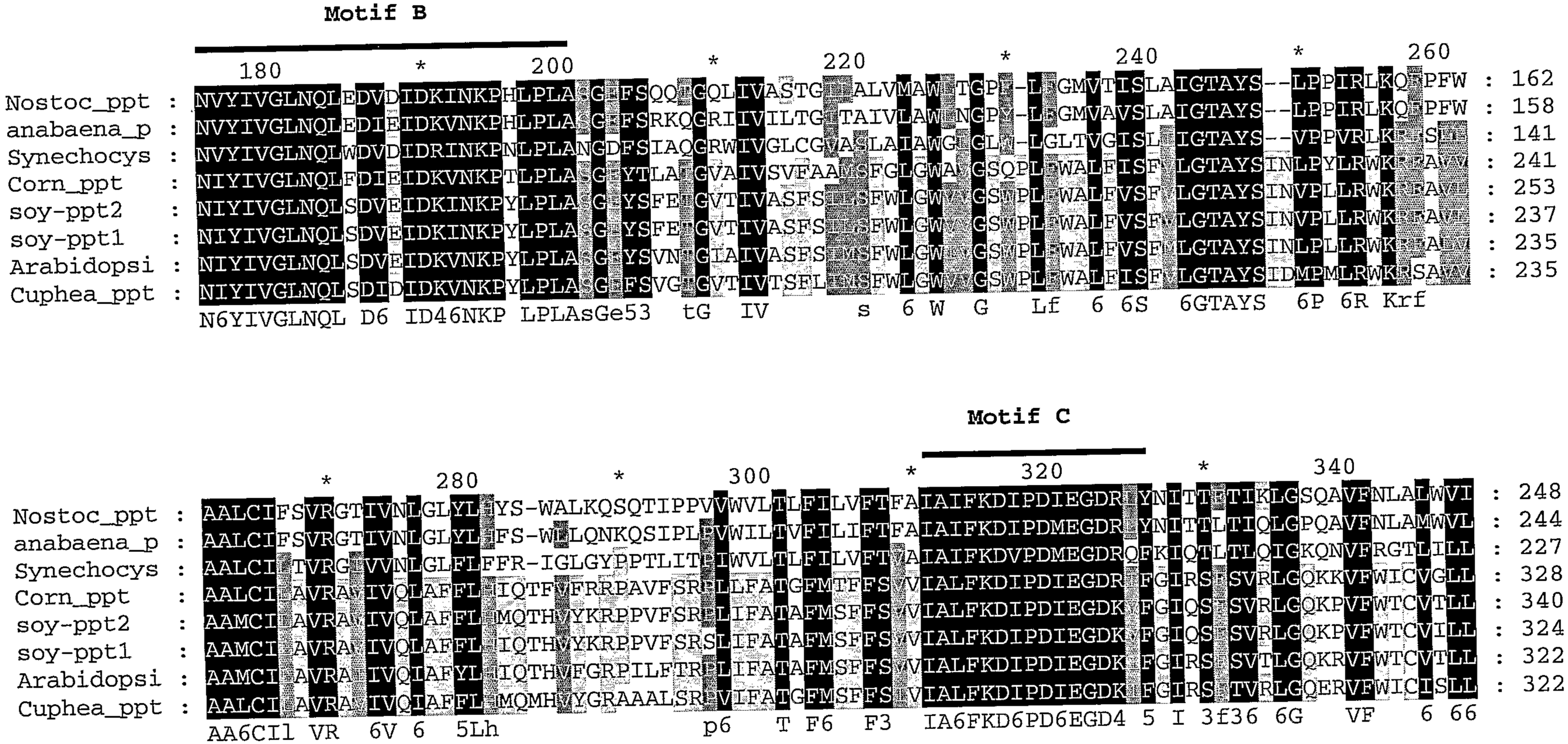
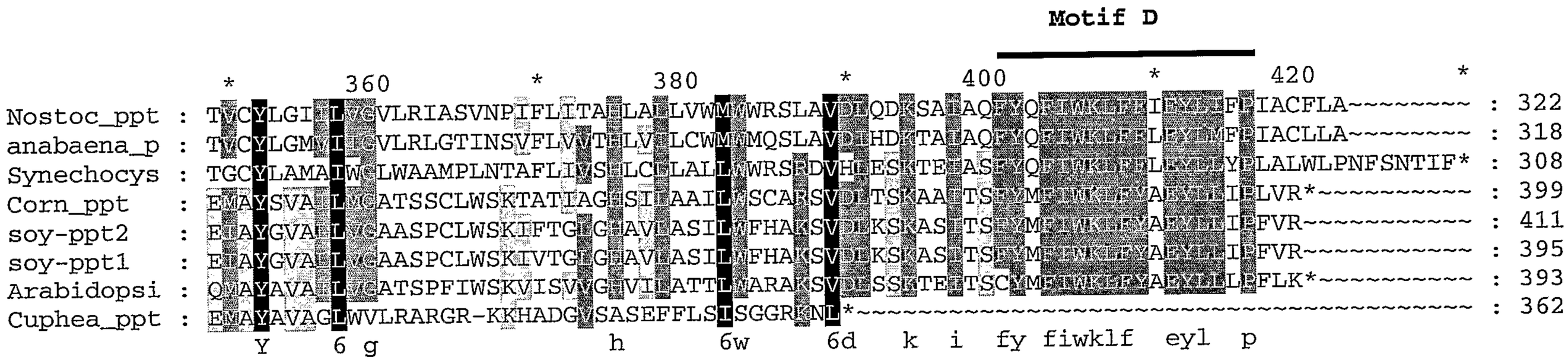

Motif B
Motif C
Nostoc_ppt
anabaena_p
Synechocys
Corn_ppt
soy-ppt2
soy-ppt1
Arabidopsi
Cuphea_ppt
N6YIVGLNQL D6 ID46NKP LPLAsGe53 tG IV s 6 W G Lf 6 6S 6GTAYS 6P 6R Krf
AA6CIl VR 6V 6 5Lh p6 T F6 F3 IA6FKD6PD6EGD4 5 I 3f36 6G VF 6 66

Fig. 2b



Nostoc_ppt : (SEQ ID NO: 1)
anabaena_p : (SEQ ID NO: 2)
Synechocys : (SEQ ID NO: 3)
Corn_ppt : (SEQ ID NO: 4)
soy-ppt2 : (SEQ ID NO: 5)
soy-ppt1 : (SEQ ID NO: 6)
Arabidopsi : (SEQ ID NO: 7)
Cuphea_ppt : (SEQ ID NO: 8)

Fig. 2c

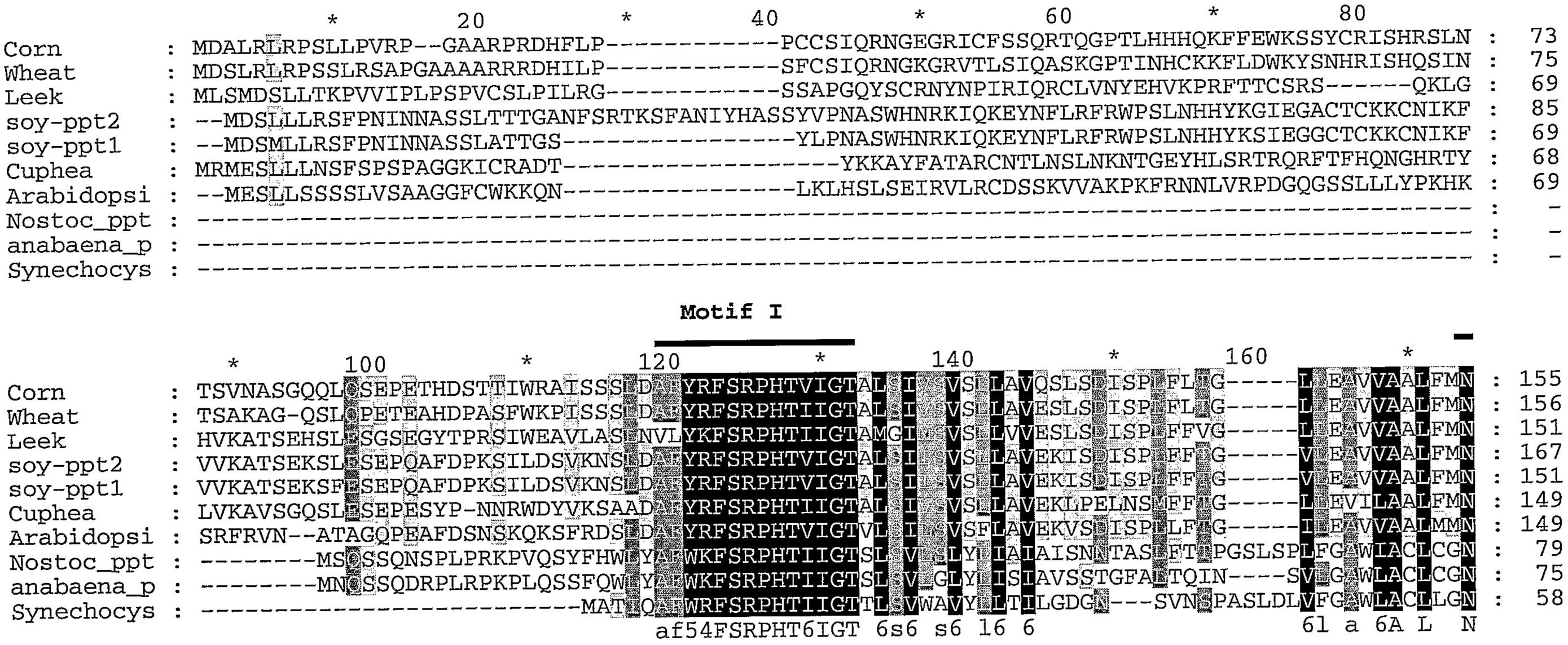

Motif I
Corn
Wheat
Leek
soy-ppt2
soy-ppt1
Cuphea
Arabidopsi
Nostoc_ppt
anabaena_p
Synechocys

Fig. 3a

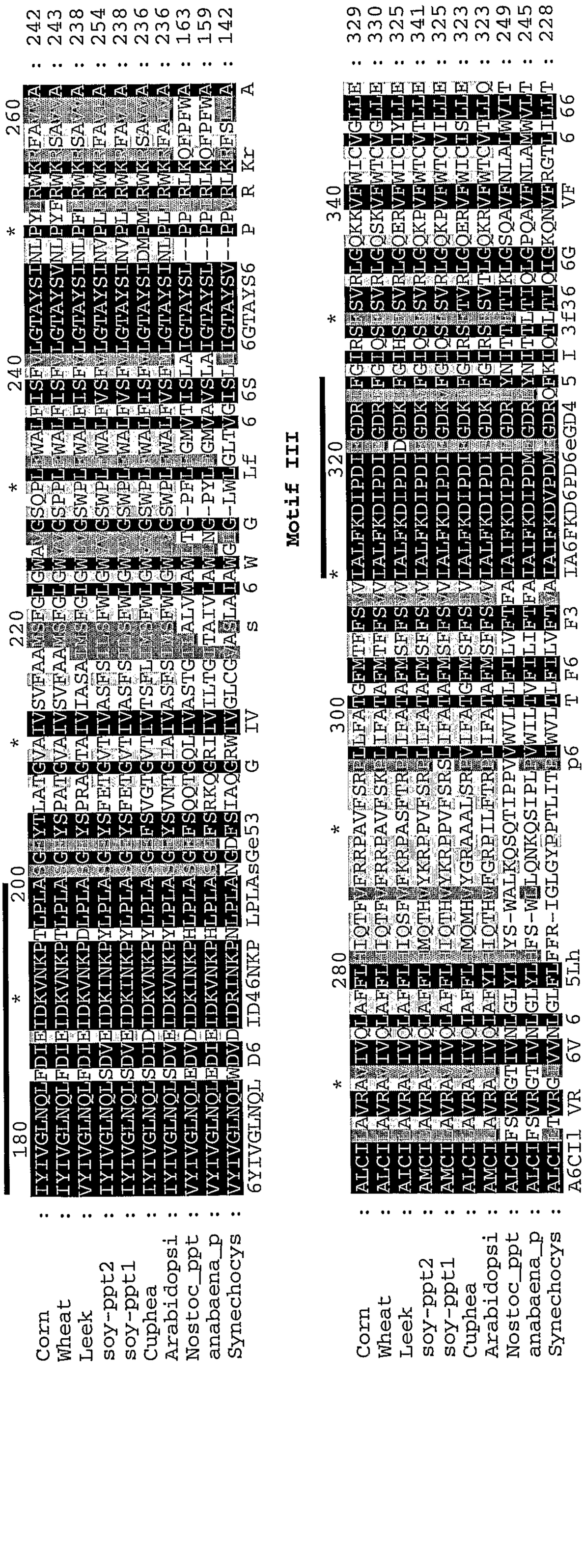
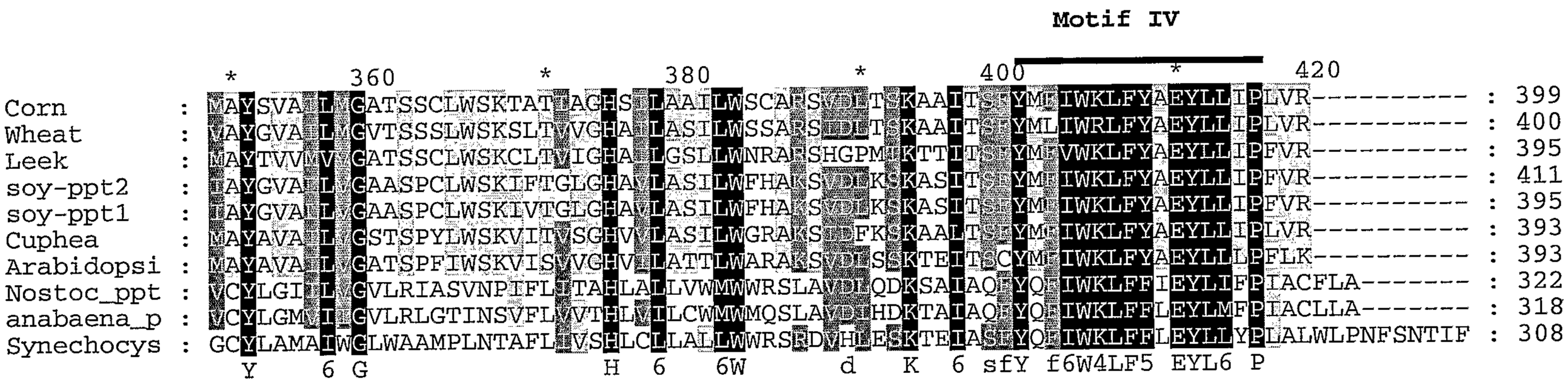

Motif II
Motif III
Corn
Wheat
Leek
soy-ppt2
soy-ppt1
Cuphea
Arabidopsi
Nostoc_ppt
anabaena_p
Synechocys

Fig. 3b



Corn : (SEQ ID NO: 4)
Wheat : (SEQ ID NO: 10)
Leek : (SEQ ID NO: 9)
soy-ppt2 : (SEQ ID NO: 5)
soy-ppt1 : (SEQ ID NO: 6)
Cuphea : (SEQ ID NO: 11)
Arabidopsi : (SEQ ID NO: 7)
Nostoc_ppt : (SEQ ID NO: 1)
anabaena_p : (SEQ ID NO: 2)
Synechocys : (SEQ ID NO: 3)

Fig. 3c

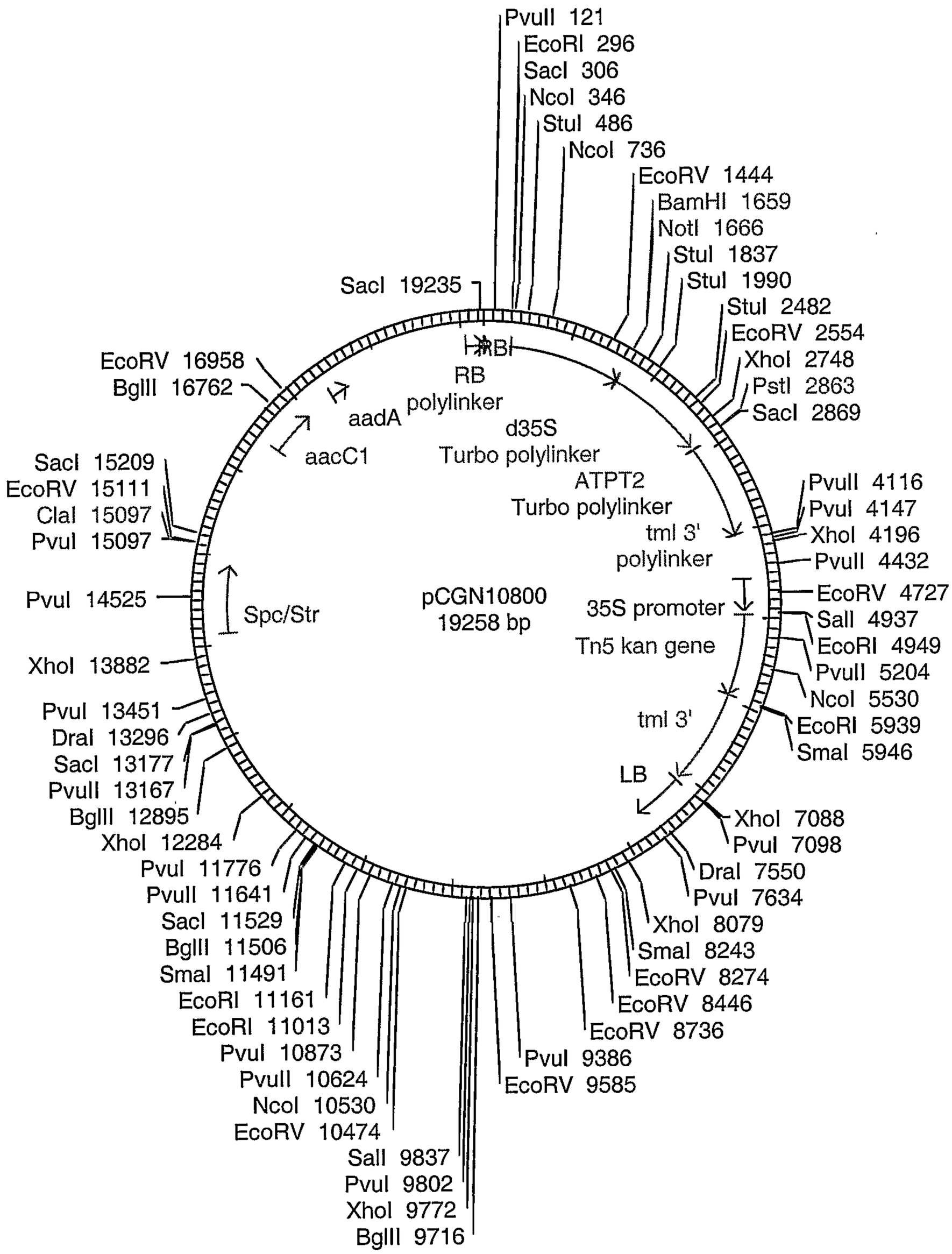

PvuII 121
EcoRI 296
SacI 306
NcoI 346
StuI 486
NcoI 736
EcoRV 1444
BamHI 1659
NotI 1666
StuI 1837
StuI 1990
StuI 2482
EcoRV 2554
XhoI 2748
PstI 2863
SacI 2869
SacI 19235
RB
RB
polylinker
d35S
Turbo polylinker
ATPT2
Turbo polylinker
tml 3'
polylinker
35S promoter
Tn5 kan gene
tml 3'
LB
aadA
aacC1
Spc/Str
pCGN10800
19258 bp
EcoRV 16958
BglII 16762
SacI 15209
EcoRV 15111
ClaI 15097
PvuI 15097
PvuI 14525
XhoI 13882
PvuI 13451
DraI 13296
SacI 13177
PvuII 13167
BglII 12895
XhoI 12284
PvuI 11776
PvuII 11641
SacI 11529
BglII 11506
SmaI 11491
EcoRI 11161
EcoRI 11013
PvuI 10873
PvuII 10624
NcoI 10530
EcoRV 10474
SalI 9837
PvuI 9802
XhoI 9772
BglII 9716
PvuI 9386
EcoRV 9585
EcoRV 8736
EcoRV 8446
EcoRV 8274
SmaI 8243
XhoI 8079
DraI 7550
PvuI 7634
XhoI 7088
PvuI 7098
PvuII 4116
PvuI 4147
XhoI 4196
PvuII 4432
EcoRV 4727
SalI 4937
EcoRI 4949
PvuII 5204
NcoI 5530
EcoRI 5939
SmaI 5946

Fig. 4

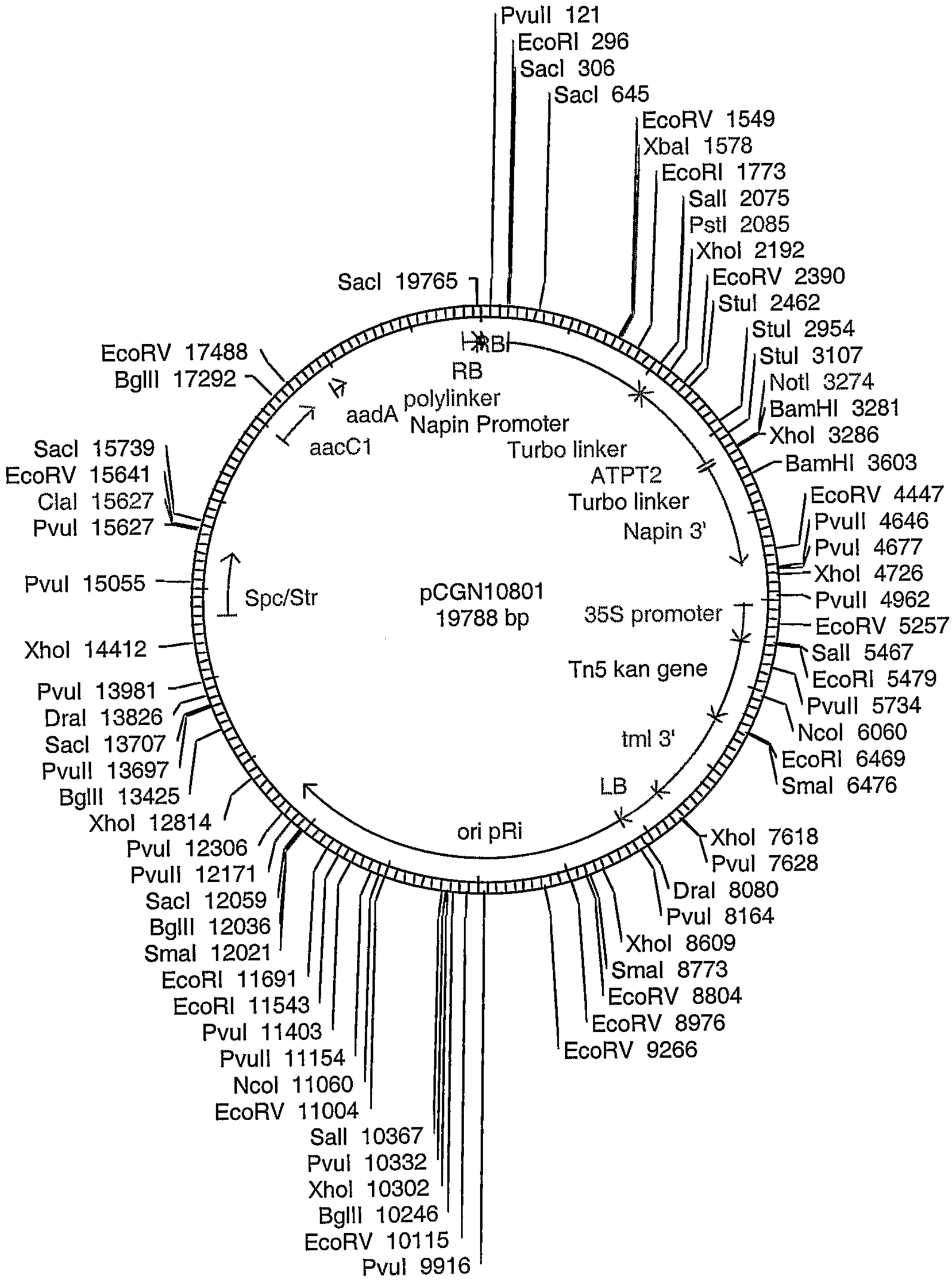

PvuII 121
EcoRI 296
SacI 306
SacI 645
EcoRV 1549
XbaI 1578
EcoRI 1773
SalI 2075
PstI 2085
XhoI 2192
EcoRV 2390
StuI 2462
StuI 2954
StuI 3107
NotI 3274
BamHI 3281
XhoI 3286
BamHI 3603
EcoRV 4447
PvuII 4646
PvuI 4677
XhoI 4726
PvuII 4962
EcoRV 5257
SalI 5467
EcoRI 5479
PvuII 5734
NcoI 6060
EcoRI 6469
SmaI 6476
XhoI 7618
PvuI 7628
DraI 8080
PvuI 8164
XhoI 8609
SmaI 8773
EcoRV 8804
EcoRV 8976
EcoRV 9266
PvuI 9916
EcoRV 10115
BglII 10246
XhoI 10302
PvuI 10332
SalI 10367
NcoI 11060
EcoRV 11004
PvuII 11154
PvuI 11403
EcoRI 11543
EcoRI 11691
SmaI 12021
BglII 12036
SacI 12059
PvuII 12171
PvuI 12306
XhoI 12814
BglII 13425
PvuII 13697
SacI 13707
DraI 13826
PvuI 13981
XhoI 14412
PvuI 15055
ClaI 15627
EcoRV 15641
SacI 15739
EcoRV 17488
BglII 17292
SacI 19765
RB
RB
polylinker
Napin Promoter
Turbo linker
ATPT2
Turbo linker
Napin 3'
35S promoter
Tn5 kan gene
tml 3'
LB
ori pRi
aadA
aacC1
Spc/Str
pCGN10801
19788 bp

Fig. 5

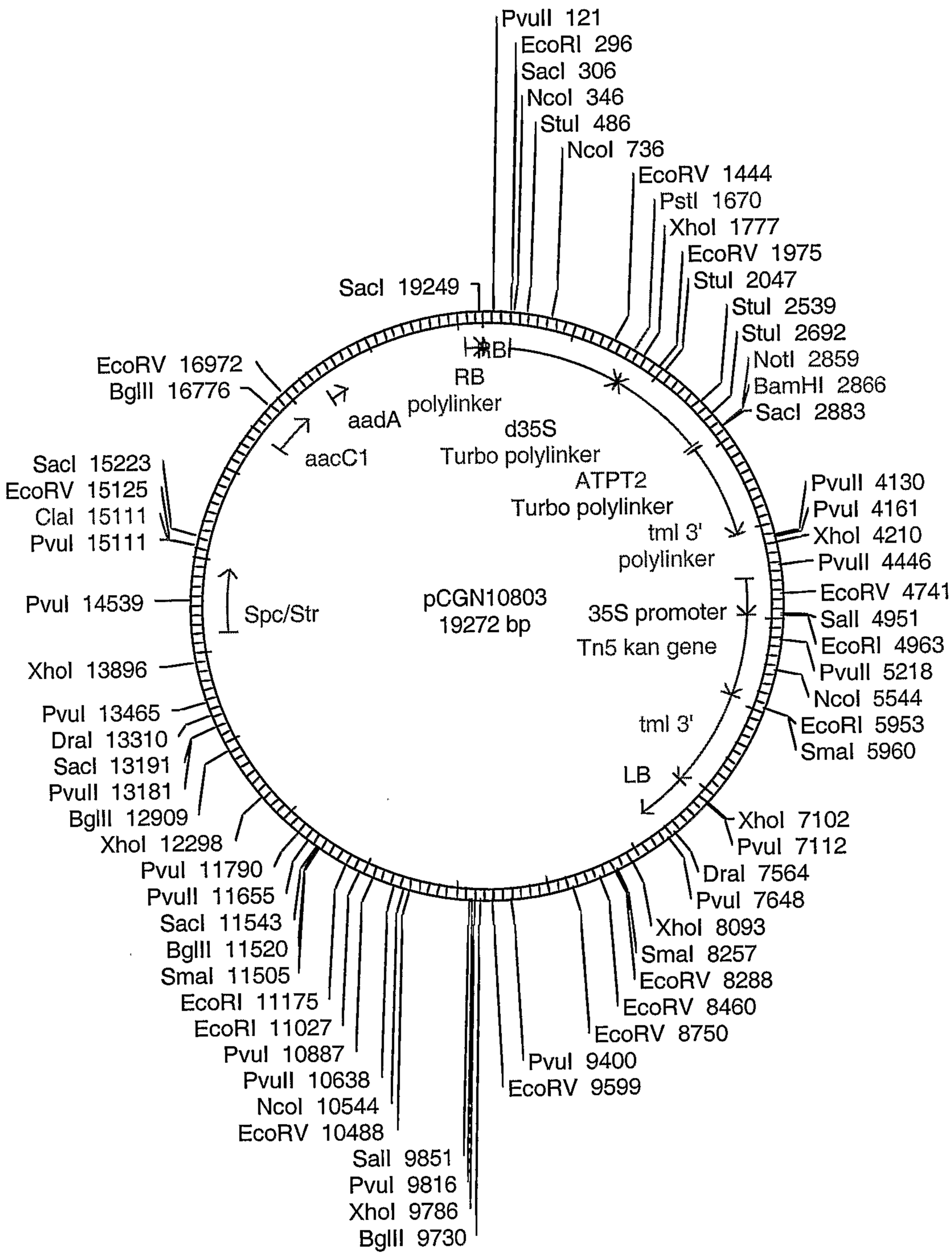

PvuII 121
EcoRI 296
SacI 306
NcoI 346
StuI 486
NcoI 736
EcoRV 1444
PstI 1670
XhoI 1777
EcoRV 1975
StuI 2047
StuI 2539
StuI 2692
NotI 2859
BamHI 2866
SacI 2883
SacI 19249
RBI
RB
polylinker
EcoRV 16972
BglII 16776
aadA
aacC1
d35S
Turbo polylinker
ATPT2
Turbo polylinker
tml 3'
polylinker
SacI 15223
EcoRV 15125
ClaI 15111
PvuI 15111
PvuII 4130
PvuI 4161
XhoI 4210
PvuII 4446
EcoRV 4741
SalI 4951
EcoRI 4963
PvuII 5218
NcoI 5544
EcoRI 5953
SmaI 5960
pCGN10803
19272 bp
35S promoter
Tn5 kan gene
tml 3'
LB
PvuI 14539
Spc/Str
XhoI 13896
PvuI 13465
DraI 13310
SacI 13191
PvuII 13181
BglII 12909
XhoI 12298
PvuI 11790
PvuII 11655
SacI 11543
BglII 11520
SmaI 11505
EcoRI 11175
EcoRI 11027
PvuI 10887
PvuII 10638
NcoI 10544
EcoRV 10488
SalI 9851
PvuI 9816
XhoI 9786
BglII 9730
PvuI 9400
EcoRV 9599
XhoI 7102
PvuI 7112
DraI 7564
PvuI 7648
XhoI 8093
SmaI 8257
EcoRV 8288
EcoRV 8460
EcoRV 8750

Fig. 6

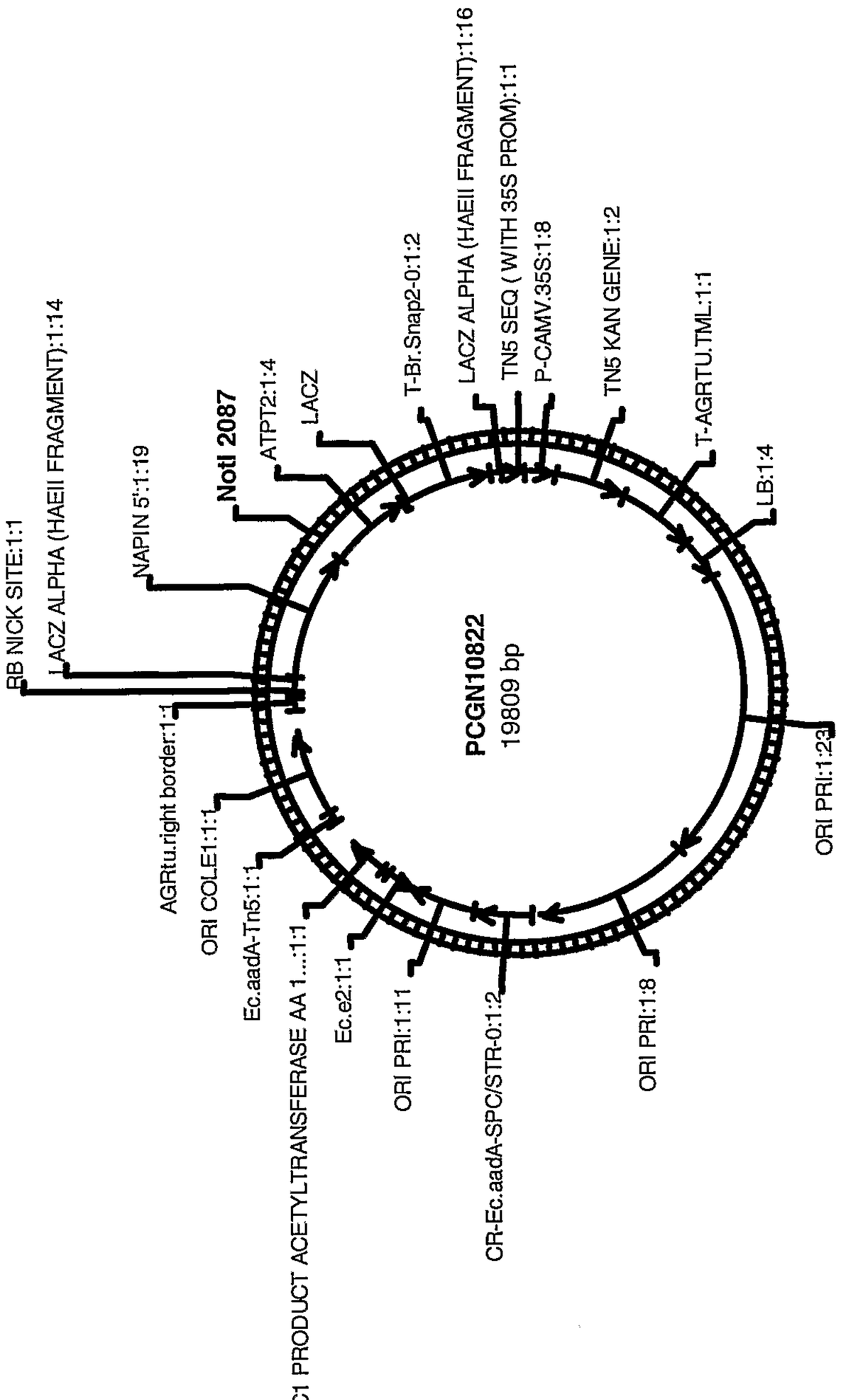

PCGN10822
19809 bp
RB NICK SITE:1:1
LACZ ALPHA (HAEII FRAGMENT):1:14
NAPIN 5':1:19
NotI 2087
ATPT2:1:4
LACZ
T-Br.Snap2-0:1:2
LACZ ALPHA (HAEII FRAGMENT):1
TN5 SEQ (WITH 35S PROM):1:1
P-CAMV.35S:1:8
TN5 KAN GENE:1:2
T-AGRTU.TML:1:1
LB:1:4
ORI PRI:1:23
AGRtu.right border:1:1
ORI COLE1:1:1
Ec.aadA-Tn5:1:1
AACC1 PRODUCT ACETYLTRANSFERASE AA 1...:1:1
Ec.e2:1:1
ORI PRI:1:11
CR-Ec.aadA-SPC/STR-0:1:2
ORI PRI:1:8

Fig. 7

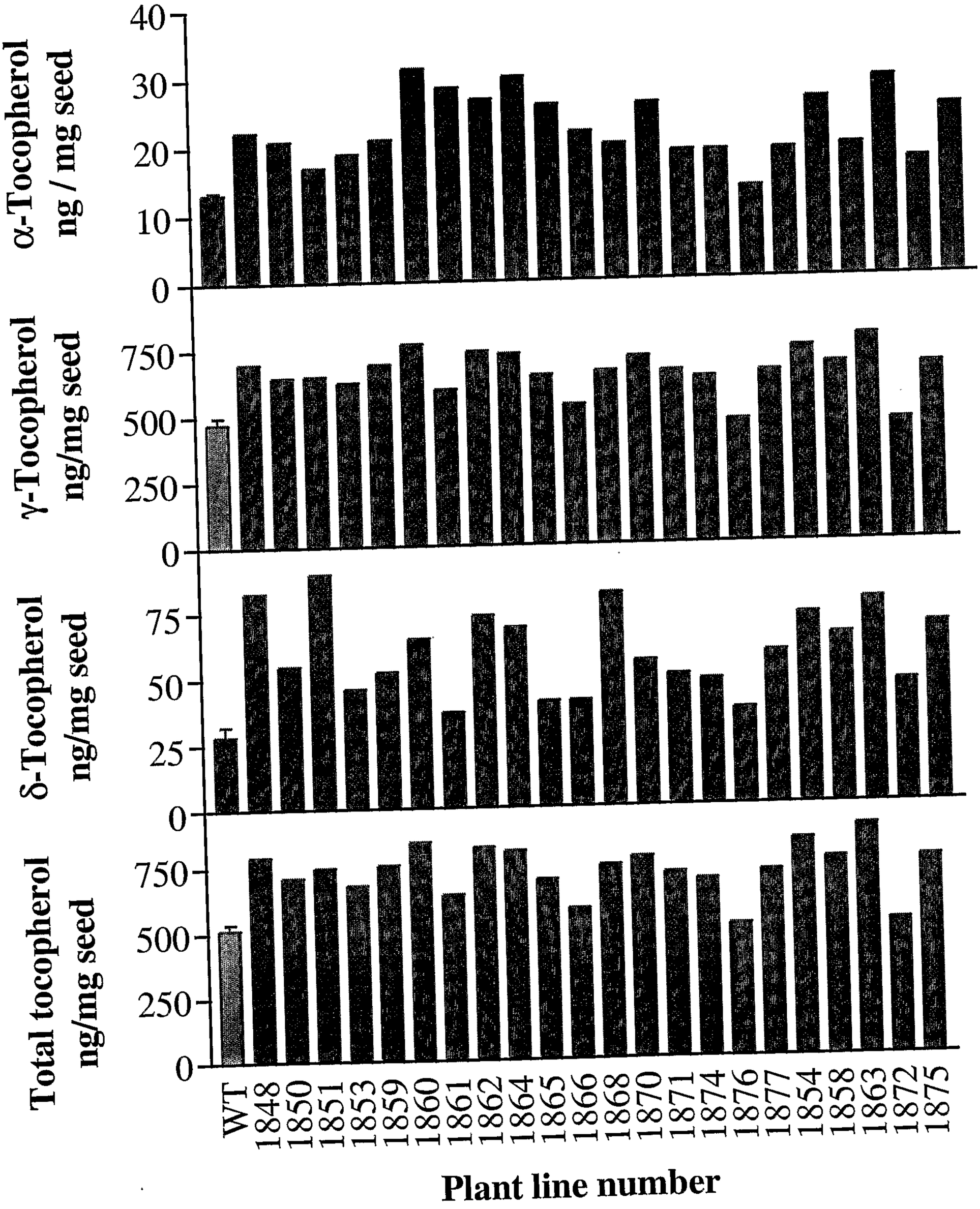

α-Tocopherol ng / mg seed
40
30
20
10
0
γ-Tocopherol ng/mg seed
750
500
250
0
δ-Tocopherol ng/mg seed
75
50
25
0
Total tocopherol ng/mg seed
750
500
250
0
WT
1848
1850
1851
1853
1859
1860
1861
1862
1864
1865
1866
1868
1870
1871
1874
1876
1877
1854
1858
1863
1872
1875
Plant line number

Fig. 8

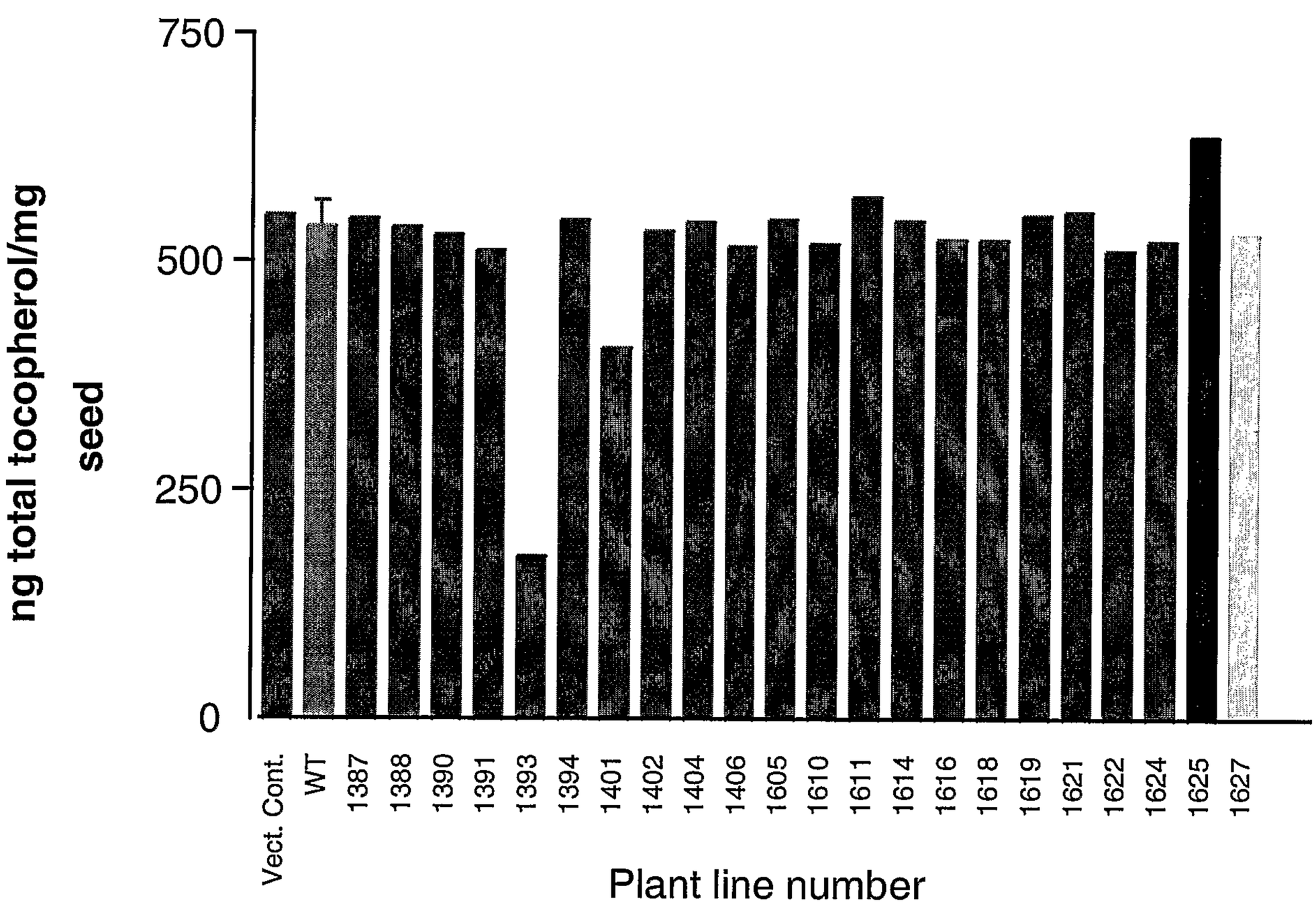

ng total tocopherol/mg seed
750
500
250
0
Vect. Cont.
WT
1387
1388
1390
1391
1393
1394
1401
1402
1404
1406
1605
1610
1611
1614
1616
1618
1619
1621
1622
1624
1625
1627
Plant line number

Fig. 9

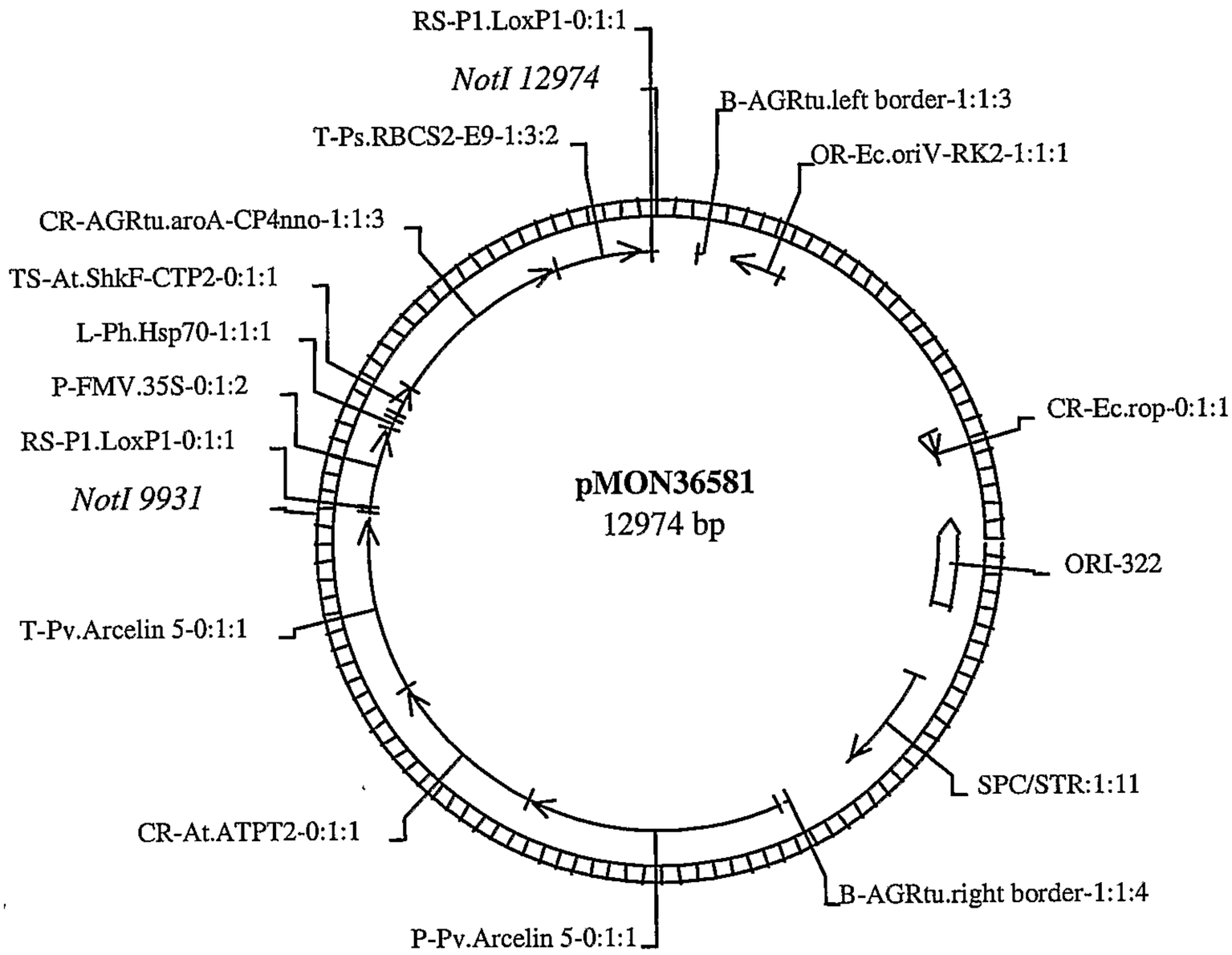

RS-P1.LoxP1-0:1:1
NotI 12974
B-AGRtu.left border-1:1:3
T-Ps.RBCS2-E9-1:3:2
OR-Ec.oriV-RK2-1:1:1
CR-AGRtu.aroA-CP4nno-1:1:3
TS-At.ShkF-CTP2-0:1:1
L-Ph.Hsp70-1:1:1
P-FMV.35S-0:1:2
RS-P1.LoxP1-0:1:1
NotI 9931
CR-Ec.rop-0:1:1
pMON36581
12974 bp
ORI-322
T-Pv.Arcelin 5-0:1:1
SPC/STR:1:11
CR-At.ATPT2-0:1:1
B-AGRtu.right border-1:1:4
P-Pv.Arcelin 5-0:1:1

Fig. 10

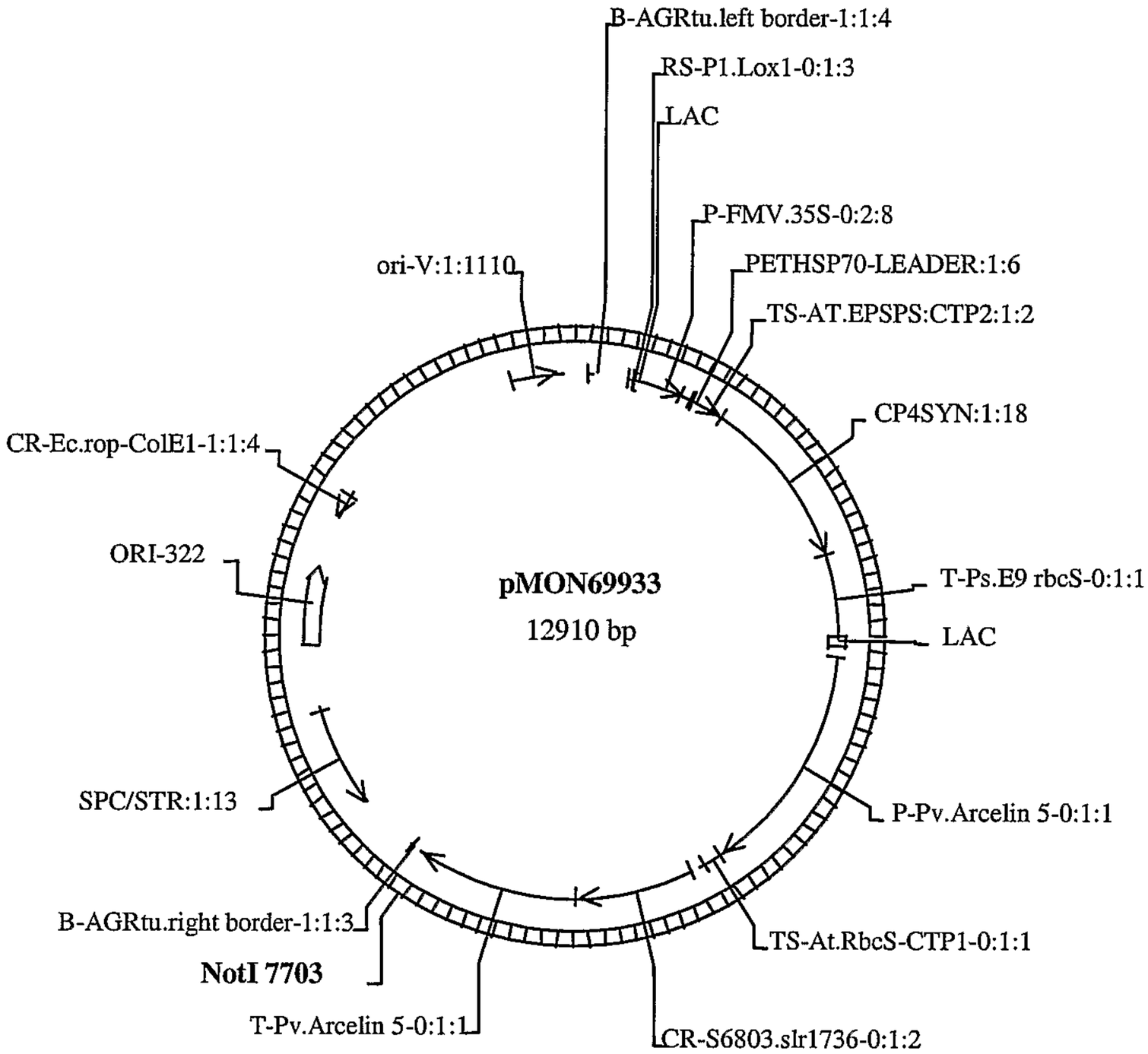

B-AGRtu.left border-1:1:4
RS-P1.Lox1-0:1:3
LAC
P-FMV.35S-0:2:8
PETHSP70-LEADER:1:6
TS-AT.EPSPS:CTP2:1:2
CP4SYN:1:18
T-Ps.E9 rbcS-0:1:1
LAC
P-Pv.Arcelin 5-0:1:1
TS-At.RbcS-CTP1-0:1:1
CR-S6803.slr1736-0:1:2
T-Pv.Arcelin 5-0:1:1
NotI 7703
B-AGRtu.right border-1:1:3
SPC/STR:1:13
ORI-322
CR-Ec.rop-ColE1-1:1:4
ori-V:1:1110
pMON69933
12910 bp

Fig. 11

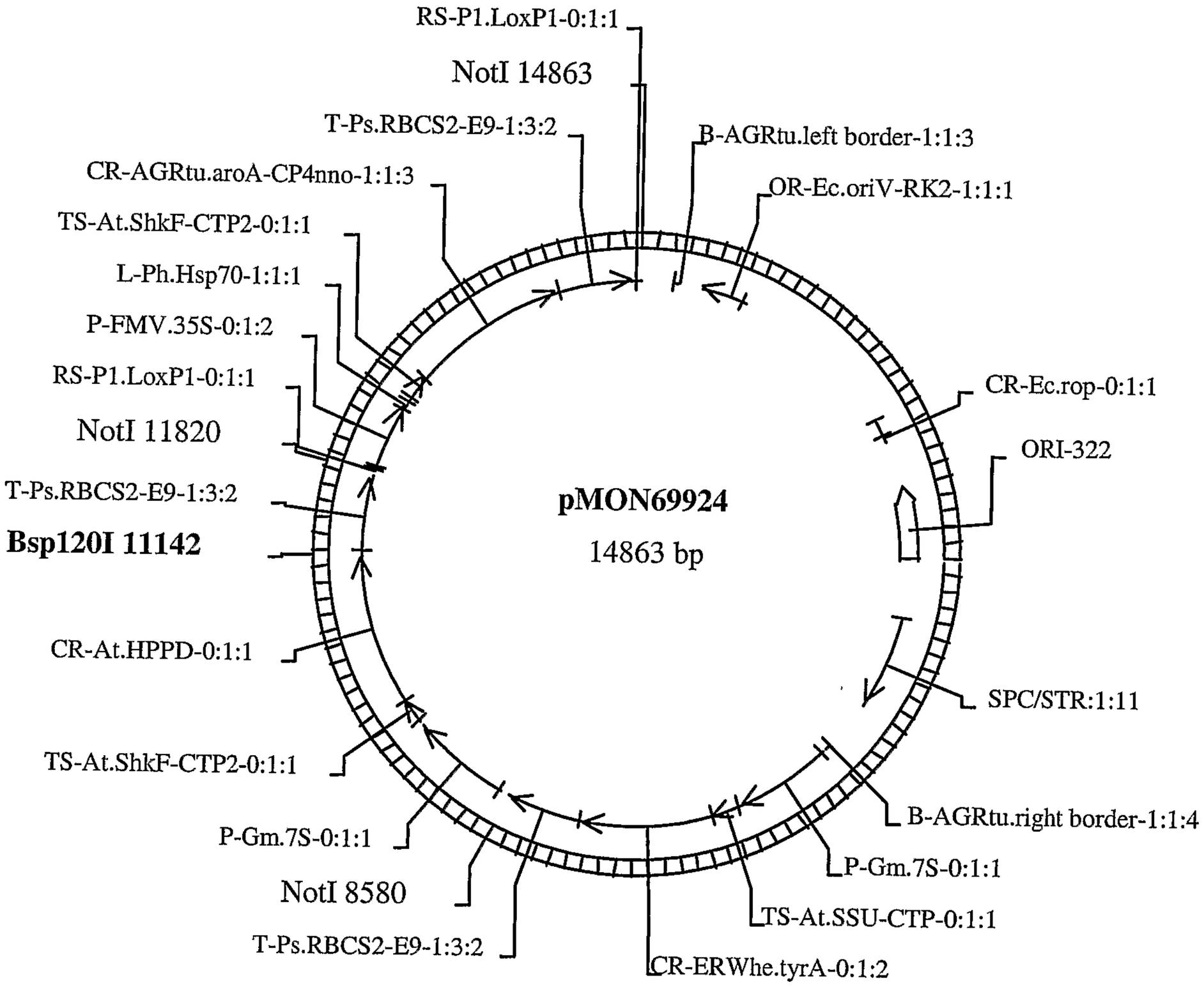

RS-P1.LoxP1-0:1:1
NotI 14863
T-Ps.RBCS2-E9-1:3:2
B-AGRtu.left border-1:1:3
CR-AGRtu.aroA-CP4nno-1:1:3
OR-Ec.oriV-RK2-1:1:1
TS-At.ShkF-CTP2-0:1:1
L-Ph.Hsp70-1:1:1
P-FMV.35S-0:1:2
RS-P1.LoxP1-0:1:1
CR-Ec.rop-0:1:1
NotI 11820
ORI-322
T-Ps.RBCS2-E9-1:3:2
pMON69924
Bsp120I 11142
14863 bp
CR-At.HPPD-0:1:1
SPC/STR:1:11
TS-At.ShkF-CTP2-0:1:1
B-AGRtu.right border-1:1:4
P-Gm.7S-0:1:1
P-Gm.7S-0:1:1
NotI 8580
TS-At.SSU-CTP-0:1:1
T-Ps.RBCS2-E9-1:3:2
CR-ERWhe.tyrA-0:1:2

Fig. 12

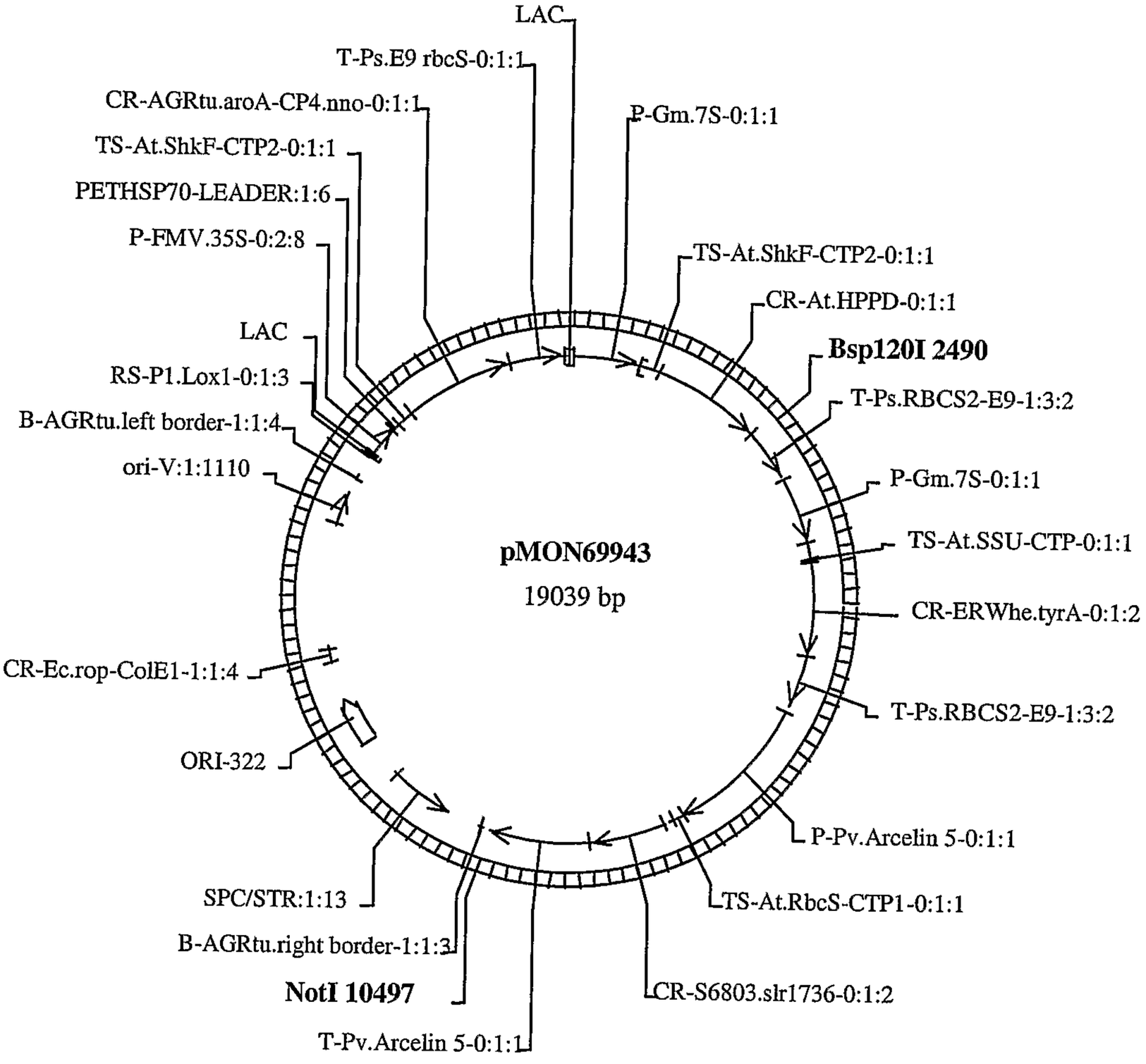

LAC
T-Ps.E9 rbcS-0:1:1
CR-AGRtu.aroA-CP4.nno-0:1:1
P-Gm.7S-0:1:1
TS-At.ShkF-CTP2-0:1:1
PETHSP70-LEADER:1:6
P-FMV.35S-0:2:8
TS-At.ShkF-CTP2-0:1:1
CR-At.HPPD-0:1:1
LAC
Bsp120I 2490
RS-P1.Lox1-0:1:3
T-Ps.RBCS2-E9-1:3:2
B-AGRtu.left border-1:1:4
ori-V:1:1110
P-Gm.7S-0:1:1
pMON69943
19039 bp
TS-At.SSU-CTP-0:1:1
CR-ERWhe.tyrA-0:1:2
CR-Ec.rop-ColE1-1:1:4
T-Ps.RBCS2-E9-1:3:2
ORI-322
P-Pv.Arcelin 5-0:1:1
SPC/STR:1:13
TS-At.RbcS-CTP1-0:1:1
B-AGRtu.right border-1:1:3
CR-S6803.slr1736-0:1:2
NotI 10497
T-Pv.Arcelin 5-0:1:1

Fig. 13

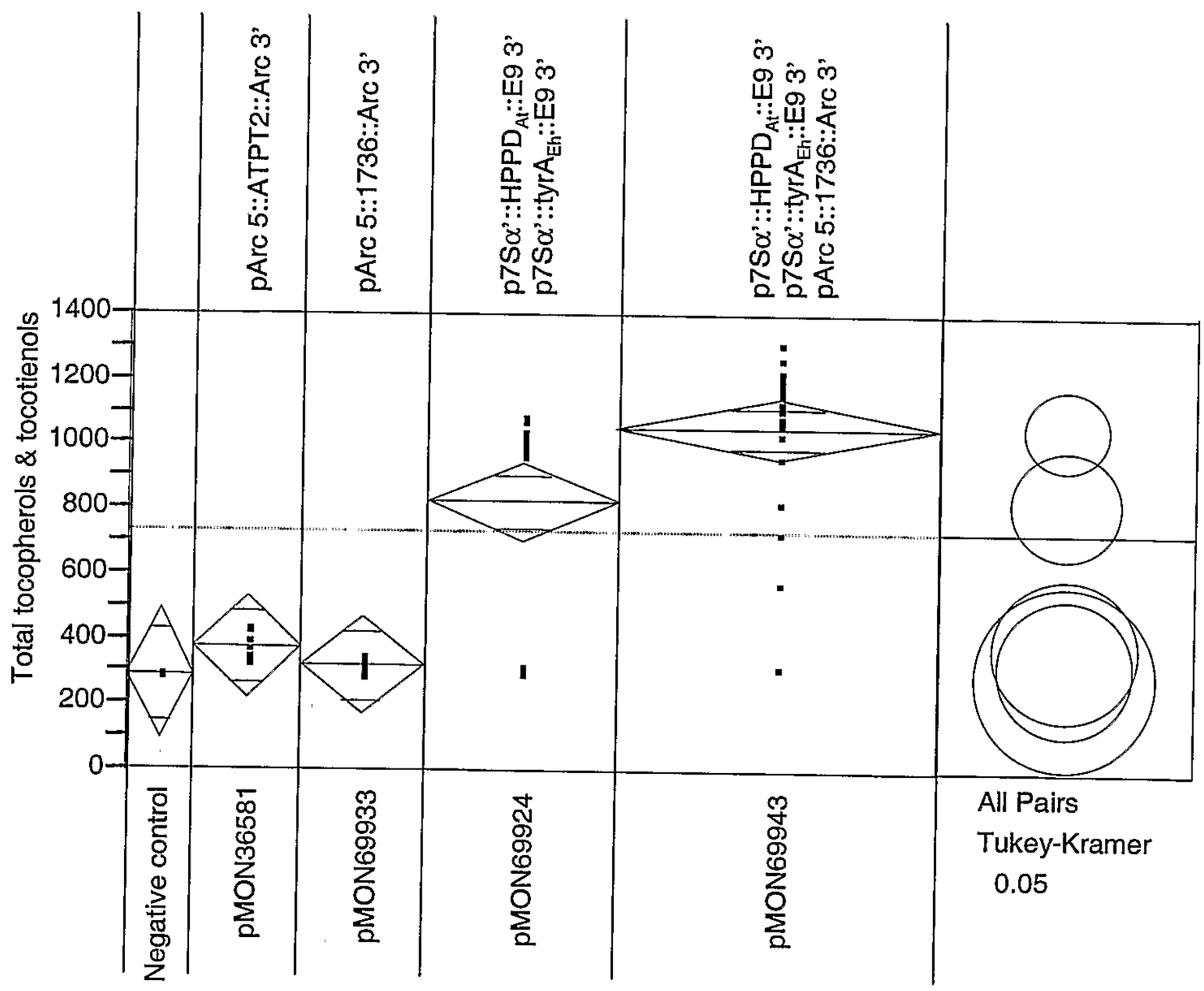

Total tocopherols & tocotienols
1400
1200
1000
800
600
400
200
0
pArc 5::ATPT2::Arc 3'
pArc 5::1736::Arc 3'
p7Sα'::HPPD$_{At}$::E9 3'
p7Sα'::tyrA$_{Eh}$::E9 3'
p7Sα'::HPPD$_{At}$::E9 3'
p7Sα'::tyrA$_{Eh}$::E9 3'
pArc 5::1736::Arc 3'
Negative control
pMON36581
pMON69933
pMON69924
pMON69943
All Pairs
Tukey-Kramer
0.05
Construct / line

Fig. 14

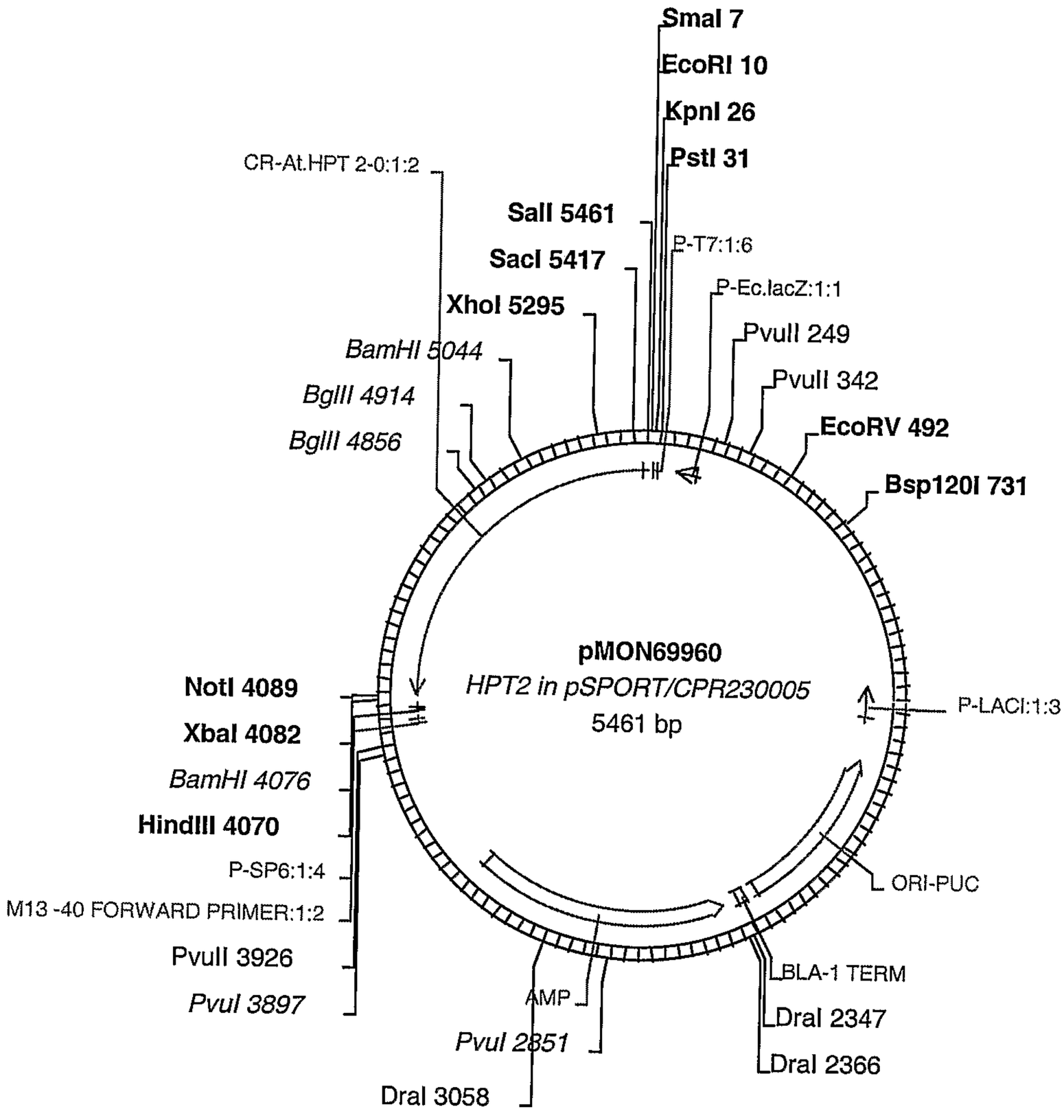

SmaI 7
EcoRI 10
KpnI 26
PstI 31
SalI 5461
SacI 5417
XhoI 5295
BamHI 5044
BglII 4914
BglII 4856
CR-At.HPT 2-0:1:2
P-T7:1:6
P-Ec.lacZ:1:1
PvuII 249
PvuII 342
EcoRV 492
Bsp120I 731
pMON69960
HPT2 in pSPORT/CPR230005
5461 bp
P-LACI:1:3
ORI-PUC
BLA-1 TERM
DraI 2347
DraI 2366
AMP
PvuI 2851
DraI 3058
NotI 4089
XbaI 4082
BamHI 4076
HindIII 4070
P-SP6:1:4
M13 -40 FORWARD PRIMER:1:2
PvuII 3926
PvuI 3897

Fig. 15

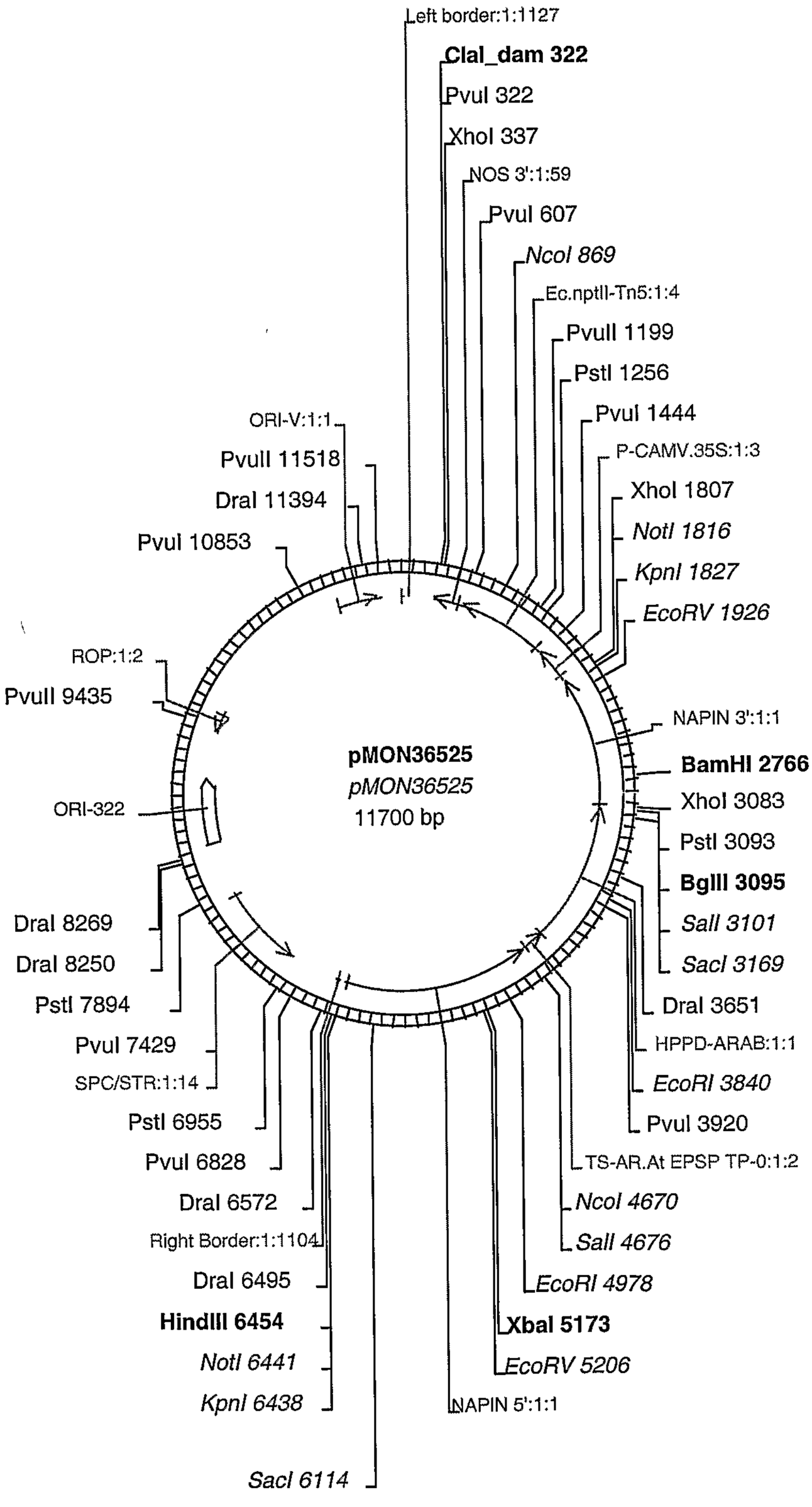

Left border:1:1127
ClaI_dam 322
PvuI 322
XhoI 337
NOS 3':1:59
PvuI 607
NcoI 869
Ec.nptII-Tn5:1:4
PvuII 1199
PstI 1256
PvuI 1444
P-CAMV.35S:1:3
XhoI 1807
NotI 1816
KpnI 1827
EcoRV 1926
NAPIN 3':1:1
BamHI 2766
XhoI 3083
PstI 3093
BglII 3095
SalI 3101
SacI 3169
DraI 3651
HPPD-ARAB:1:1
EcoRI 3840
PvuI 3920
TS-AR.At EPSP TP-0:1:2
NcoI 4670
SalI 4676
EcoRI 4978
XbaI 5173
EcoRV 5206
NAPIN 5':1:1
SacI 6114
HindIII 6454
NotI 6441
KpnI 6438
DraI 6495
Right Border:1:1104
DraI 6572
PvuI 6828
PstI 6955
SPC/STR:1:14
PvuI 7429
PstI 7894
DraI 8250
DraI 8269
PvuI 9435
ROP:1:2
ORI-322
pMON36525
pMON36525
11700 bp
PvuI 10853
DraI 11394
PvuII 11518
ORI-V:1:1

Fig. 16

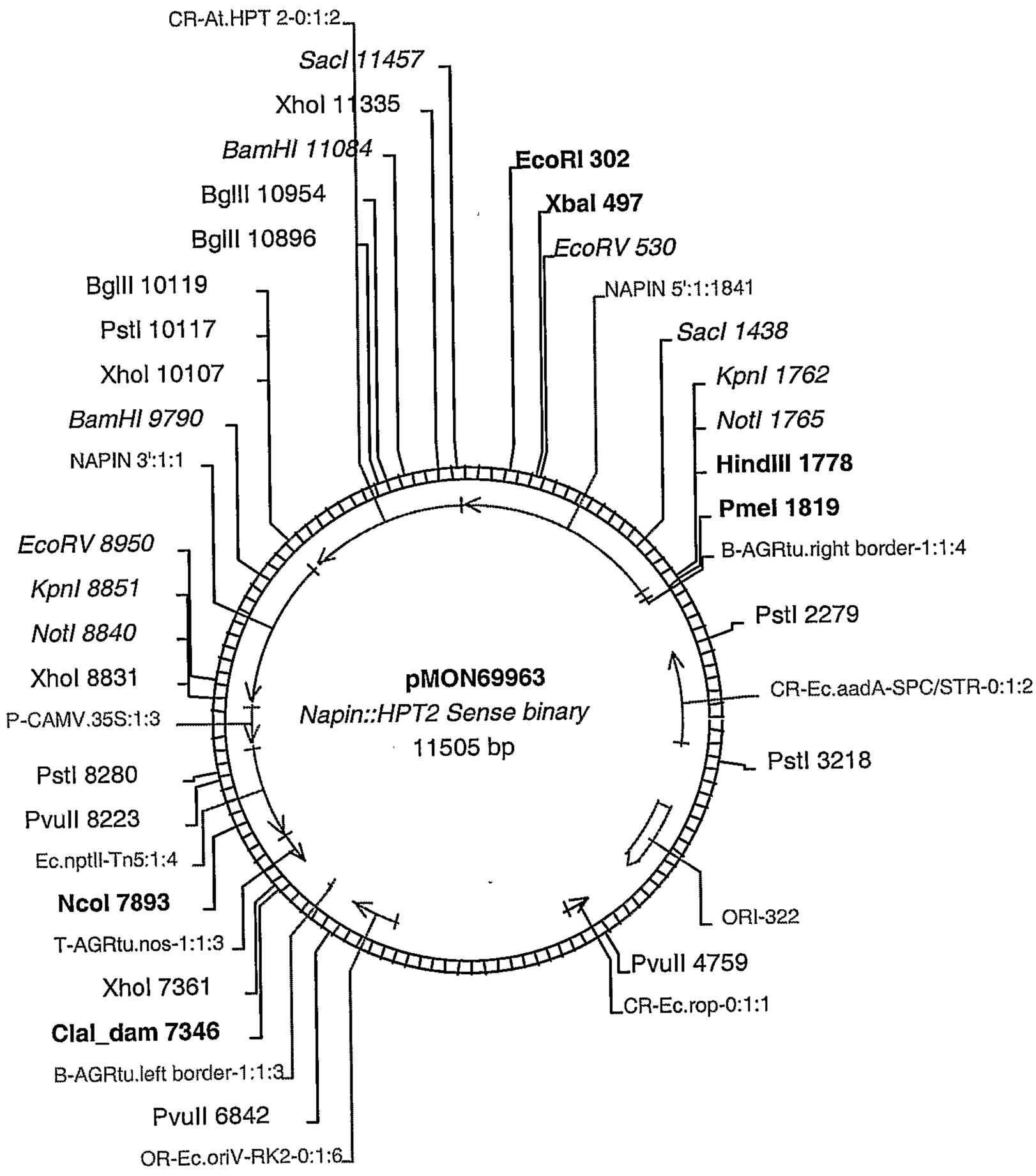

CR-At.HPT 2-0:1:2
SacI 11457
XhoI 11335
BamHI 11084
BglII 10954
BglII 10896
BglII 10119
PstI 10117
XhoI 10107
BamHI 9790
NAPIN 3':1:1
EcoRV 8950
KpnI 8851
NotI 8840
XhoI 8831
P-CAMV.35S:1:3
PstI 8280
PvuII 8223
Ec.nptII-Tn5:1:4
NcoI 7893
T-AGRtu.nos-1:1:3
XhoI 7361
ClaI_dam 7346
B-AGRtu.left border-1:1:3
PvuII 6842
OR-Ec.oriV-RK2-0:1:6
EcoRI 302
XbaI 497
EcoRV 530
NAPIN 5':1:1841
SacI 1438
KpnI 1762
NotI 1765
HindIII 1778
PmeI 1819
B-AGRtu.right border-1:1:4
PstI 2279
CR-Ec.aadA-SPC/STR-0:1:2
PstI 3218
ORI-322
PvuII 4759
CR-Ec.rop-0:1:1
pMON69963
Napin::HPT2 Sense binary
11505 bp

Fig. 17

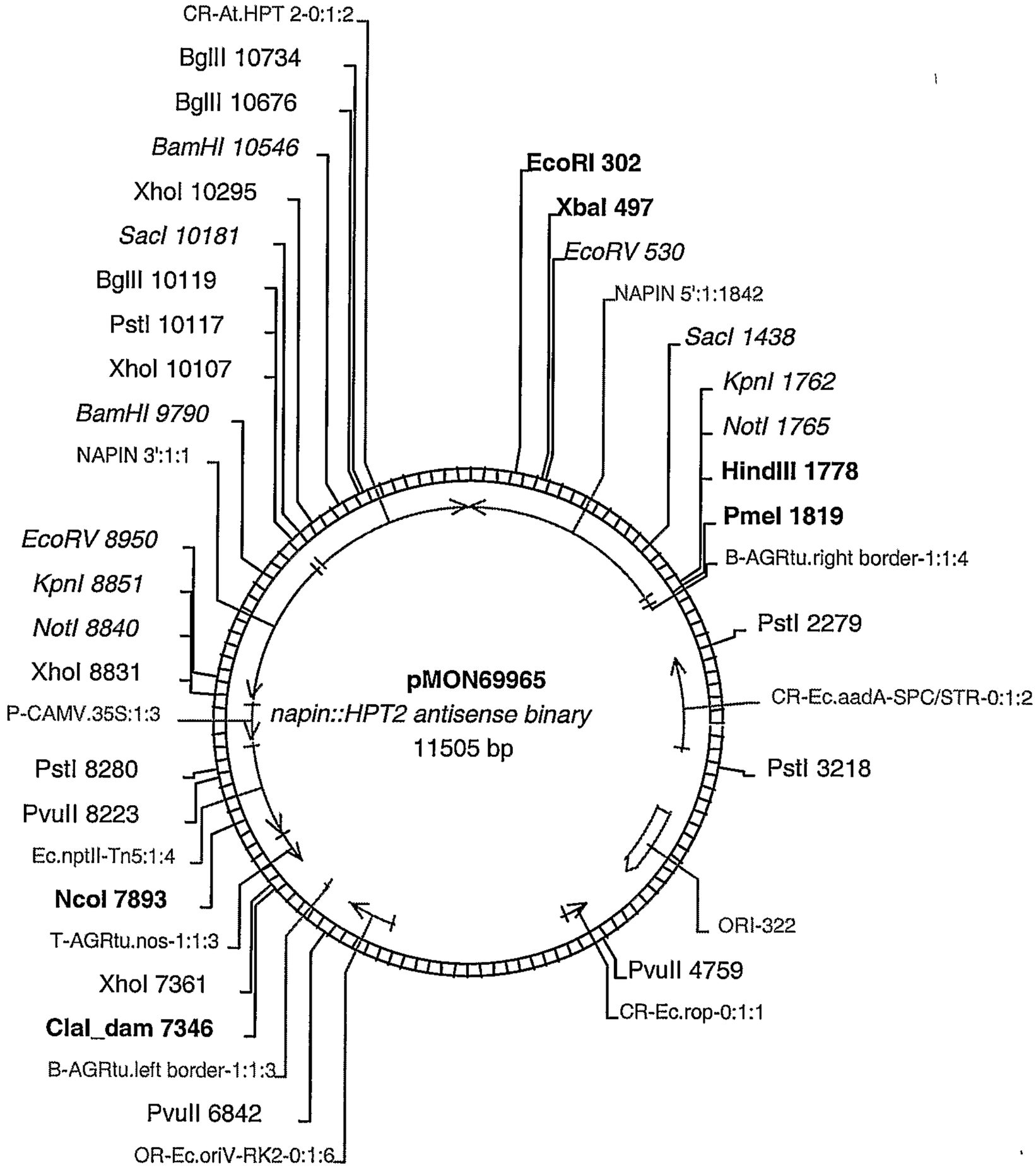

CR-At.HPT 2-0:1:2
BglII 10734
BglII 10676
BamHI 10546
XhoI 10295
SacI 10181
BglII 10119
PstI 10117
XhoI 10107
BamHI 9790
NAPIN 3':1:1
EcoRI 302
XbaI 497
EcoRV 530
NAPIN 5':1:1842
SacI 1438
KpnI 1762
NotI 1765
HindIII 1778
PmeI 1819
B-AGRtu.right border-1:1:4
PstI 2279
CR-Ec.aadA-SPC/STR-0:1:2
PstI 3218
ORI-322
PvuII 4759
CR-Ec.rop-0:1:1
EcoRV 8950
KpnI 8851
NotI 8840
XhoI 8831
P-CAMV.35S:1:3
PstI 8280
PvuII 8223
Ec.nptII-Tn5:1:4
NcoI 7893
T-AGRtu.nos-1:1:3
XhoI 7361
ClaI_dam 7346
B-AGRtu.left border-1:1:3
PvuII 6842
OR-Ec.oriV-RK2-0:1:6
pMON69965
napin::HPT2 antisense binary
11505 bp

Fig. 18

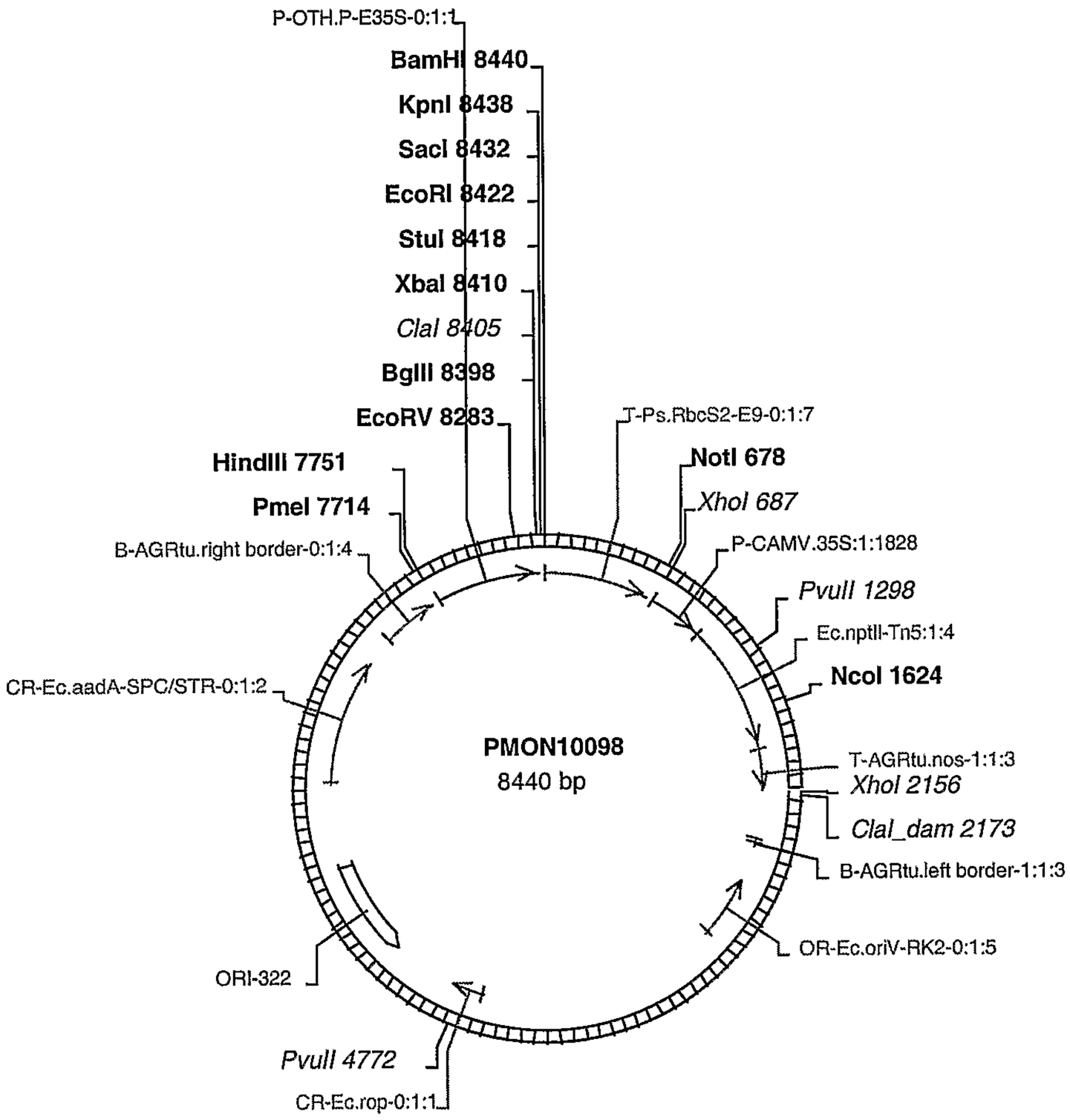

P-OTH.P-E35S-0:1:1
BamHI 8440
KpnI 8438
SacI 8432
EcoRI 8422
StuI 8418
XbaI 8410
ClaI 8405
BglII 8398
EcoRV 8283
HindIII 7751
PmeI 7714
B-AGRtu.right border-0:1:4
CR-Ec.aadA-SPC/STR-0:1:2
ORI-322
PvuII 4772
CR-Ec.rop-0:1:1
T-Ps.RbcS2-E9-0:1:7
NotI 678
XhoI 687
P-CAMV.35S:1:1828
PvuII 1298
Ec.nptII-Tn5:1:4
NcoI 1624
T-AGRtu.nos-1:1:3
XhoI 2156
ClaI_dam 2173
B-AGRtu.left border-1:1:3
OR-Ec.oriV-RK2-0:1:5
PMON10098
8440 bp

Fig. 19

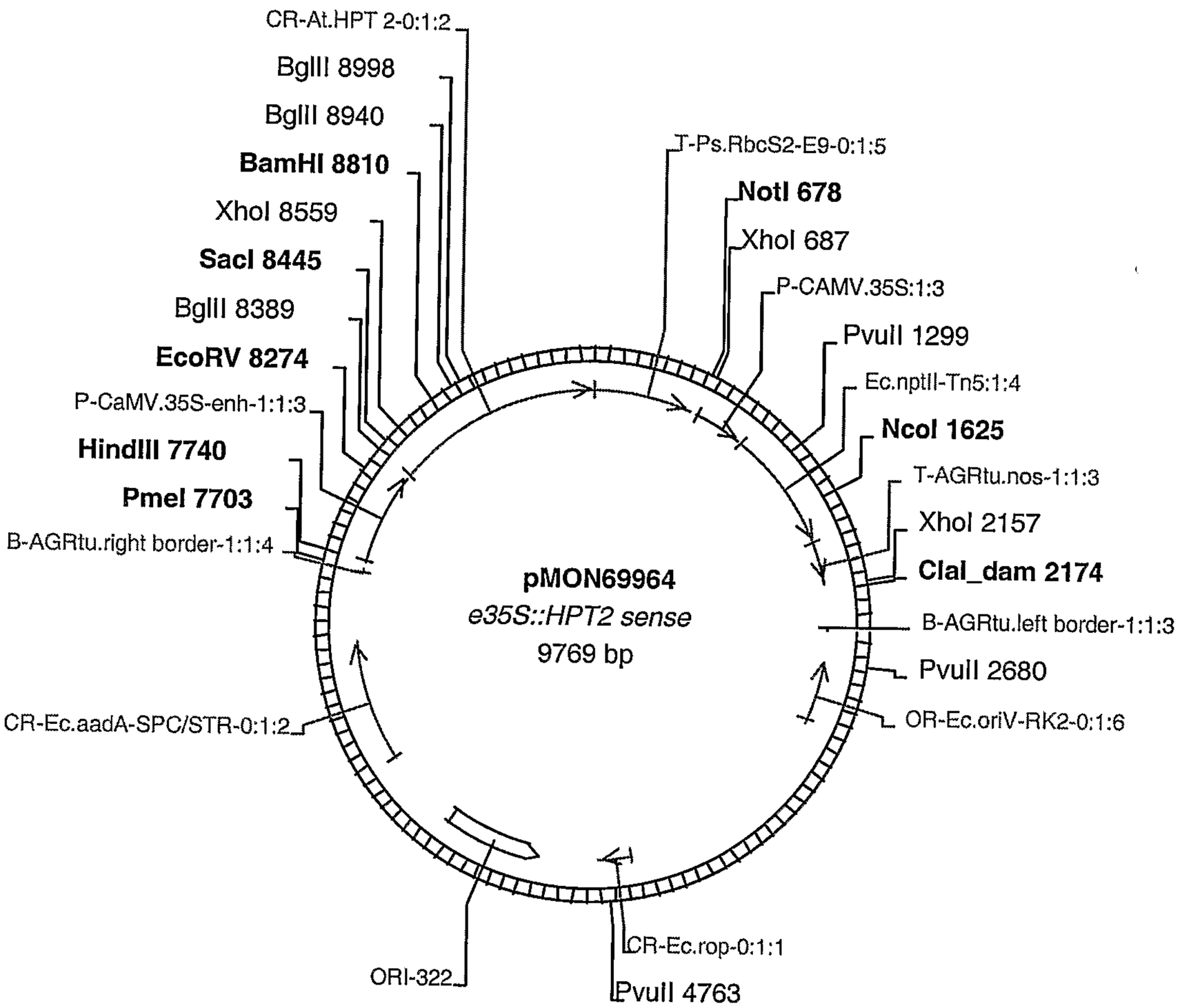

CR-At.HPT 2-0:1:2
BglII 8998
BglII 8940
BamHI 8810
XhoI 8559
SacI 8445
BglII 8389
EcoRV 8274
P-CaMV.35S-enh-1:1:3
HindIII 7740
PmeI 7703
B-AGRtu.right border-1:1:4
CR-Ec.aadA-SPC/STR-0:1:2
ORI-322
CR-Ec.rop-0:1:1
PvuII 4763
T-Ps.RbcS2-E9-0:1:5
NotI 678
XhoI 687
P-CAMV.35S:1:3
PvuII 1299
Ec.nptII-Tn5:1:4
NcoI 1625
T-AGRtu.nos-1:1:3
XhoI 2157
ClaI_dam 2174
B-AGRtu.left border-1:1:3
PvuII 2680
OR-Ec.oriV-RK2-0:1:6
pMON69964
e35S::HPT2 sense
9769 bp

Fig. 20

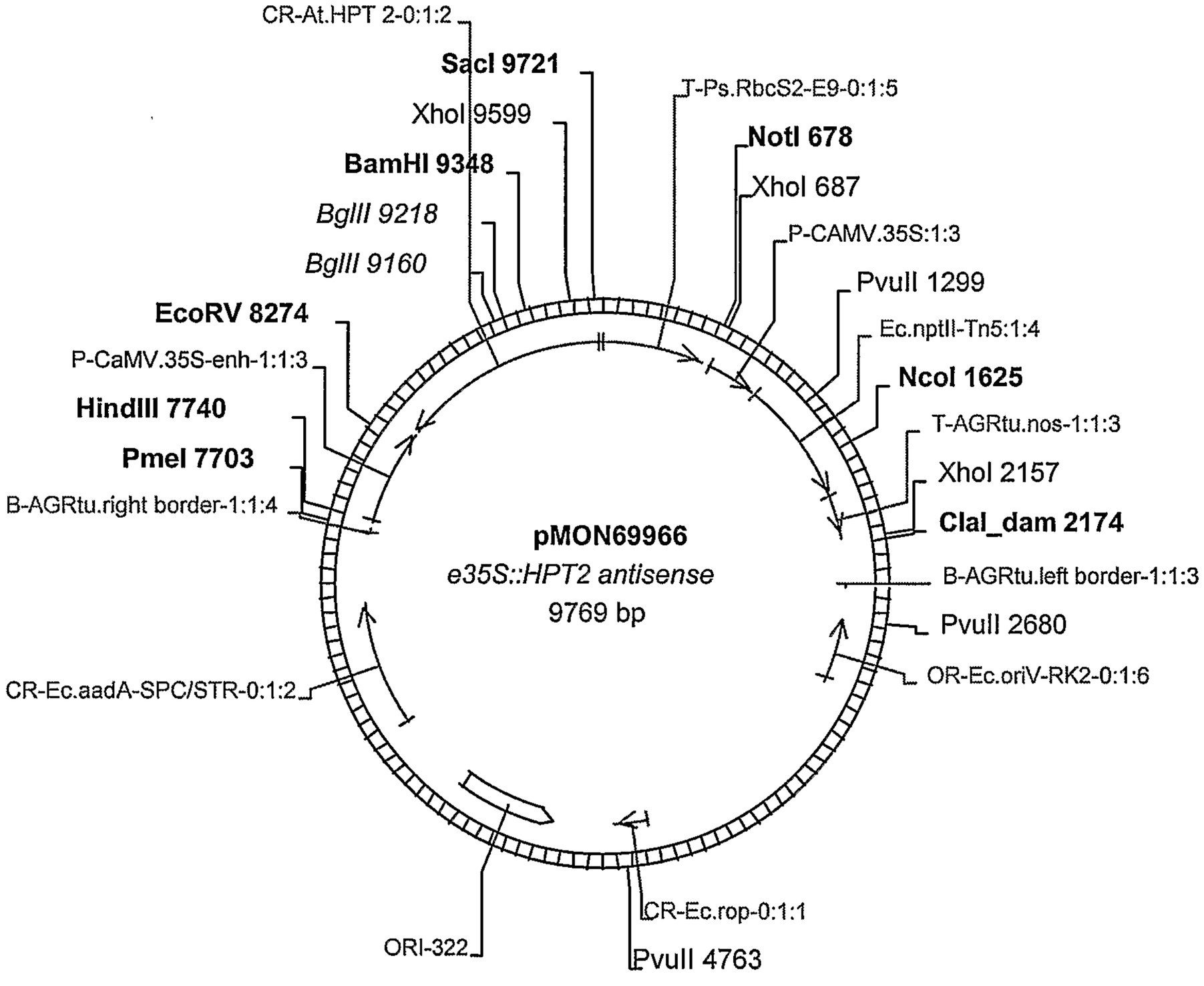

CR-At.HPT 2-0:1:2
SacI 9721
XhoI 9599
BamHI 9348
BglII 9218
BglII 9160
EcoRV 8274
P-CaMV.35S-enh-1:1:3
HindIII 7740
PmeI 7703
B-AGRtu.right border-1:1:4
T-Ps.RbcS2-E9-0:1:5
NotI 678
XhoI 687
P-CAMV.35S:1:3
PvuII 1299
Ec.nptII-Tn5:1:4
NcoI 1625
T-AGRtu.nos-1:1:3
XhoI 2157
ClaI_dam 2174
B-AGRtu.left border-1:1:3
PvuII 2680
OR-Ec.oriV-RK2-0:1:6
pMON69966
e35S::HPT2 antisense
9769 bp
CR-Ec.aadA-SPC/STR-0:1:2
ORI-322
CR-Ec.rop-0:1:1
PvuII 4763

Fig. 21

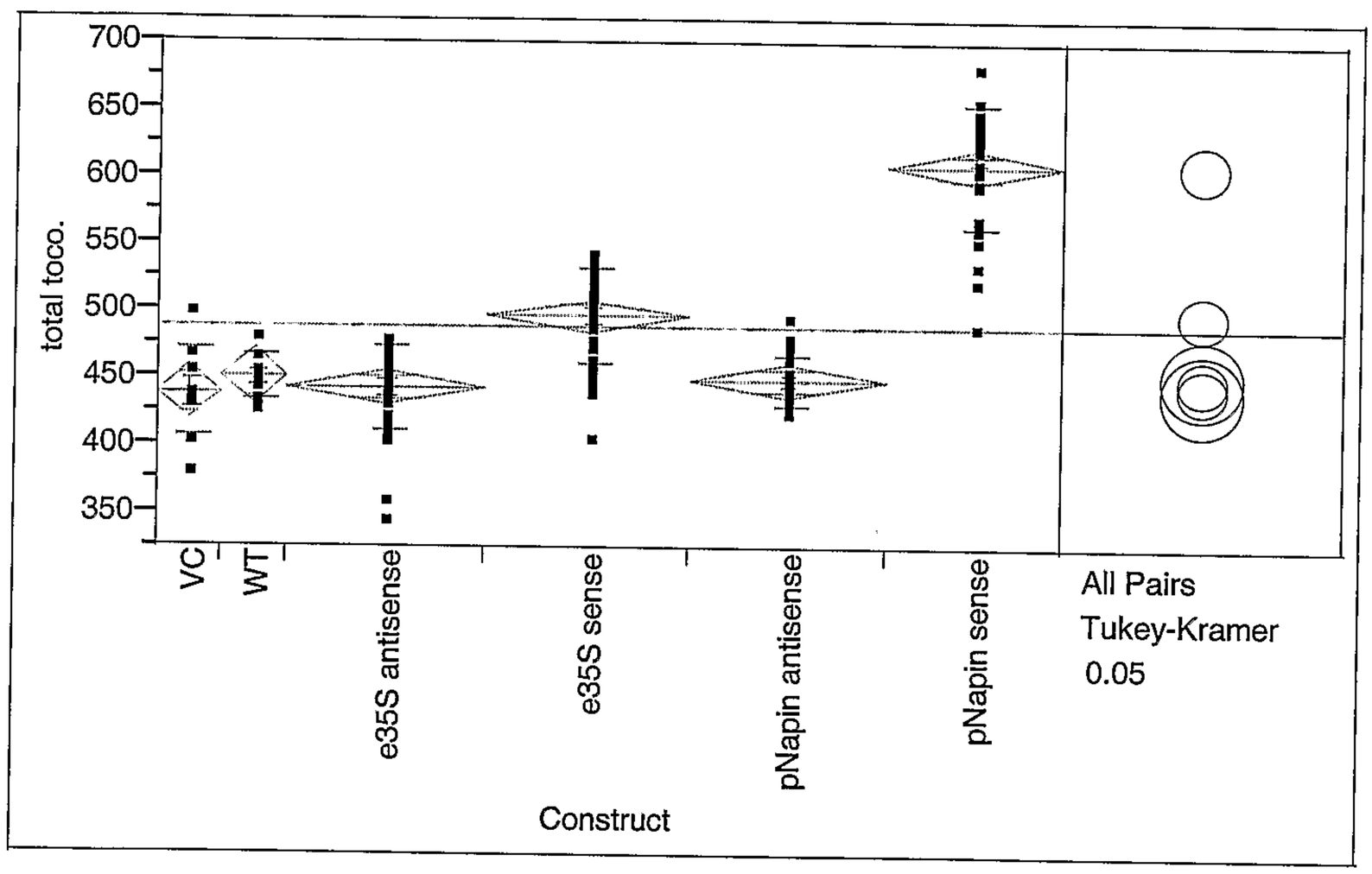

700
650
600
550
500
450
400
350
total toco.
VC
WT
e35S antisense
e35S sense
pNapin antisense
pNapin sense
All Pairs
Tukey-Kramer
0.05
Construct

Fig. 22

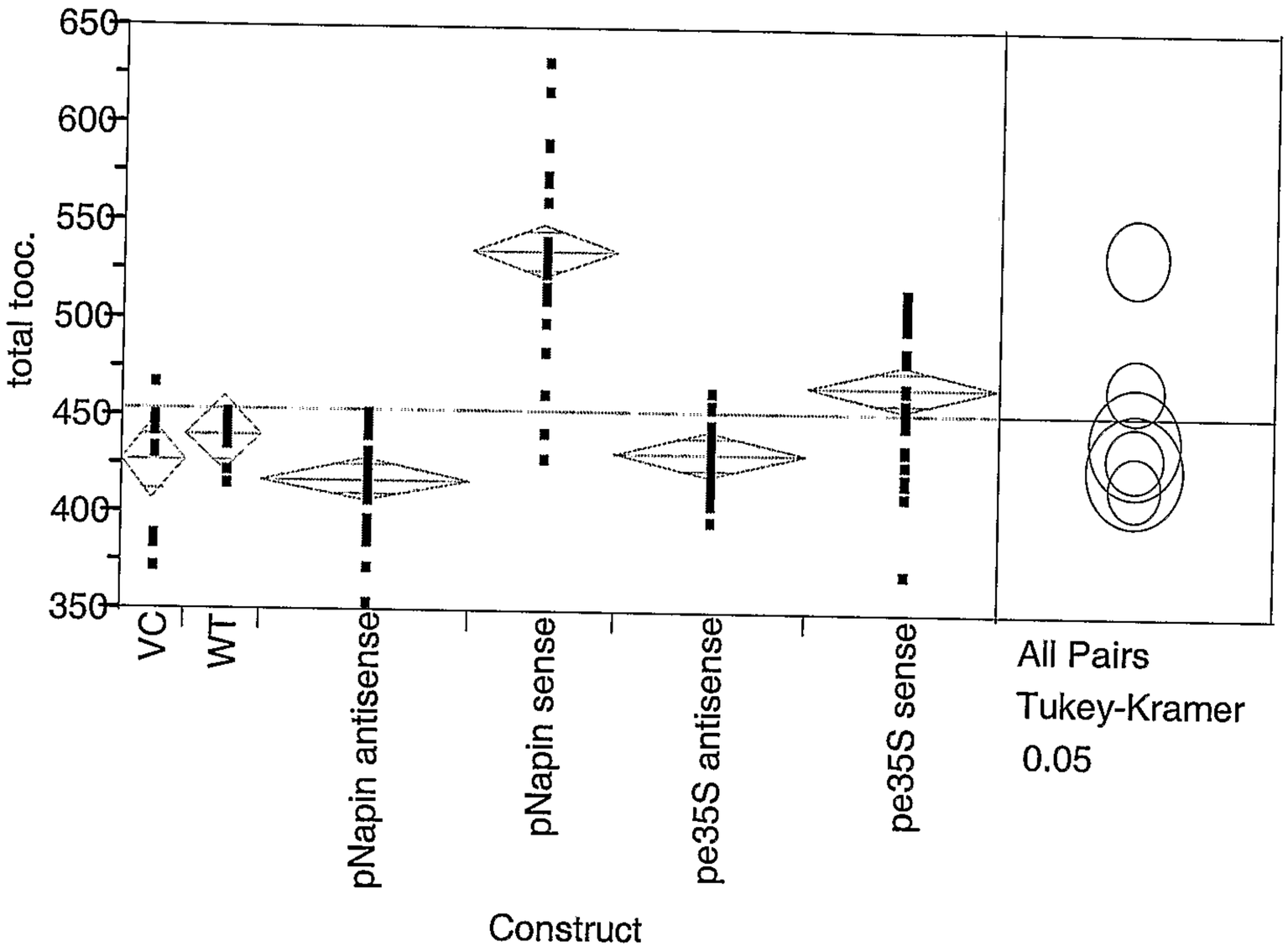

650
600
550
500
450
400
350
total tooc.
VC
WT
pNapin antisense
pNapin sense
pe35S antisense
pe35S sense
Construct
All Pairs
Tukey-Kramer
0.05

Fig. 23

```
                    *         20         *         40         *         60
NoHPT     : ------------------------------------------------------------ :   -
AnHPT     : ------------------------------------------------------------ :   -
SyHPT     : ------------------------------------------------------------ :   -
ZmHPT1    : --MDALRLRPSLLPVRPGAARPRDHFLPPCCSIQRNGEGRICFSSQRTQGPTLHHHQKFF :  58
TaHPT1    : MDSLRLRPSSLRSAPGAAAARRRDHILPSFCSIQRNGKGRVTLSIQASKGPTINHCKKFL :  60
ApHPT1    : ------MLSMDSLLTKPVVIPLPSPVCSLPILRGSSAPGQYSCRNYNPIRIQRCLVNYEH :  54
GmHPT1-1  : ------MDSMLLRSFPNINNASSLATTGSYLPNASWHNRKIQKEYNFLRFRWPSLNHHYK :  54
CpHPT1    : -------MRMESLLLNSFSPSPAGGKICRADTYKKAYFATARCNTLNSLNKNTGEYHLSR :  53
AtHPT1    : --------MESLLSSSSLVSAAGGFCWKKQNLKLHSLSEIRVLRCDSSKVVAKPKFRNNL :  52
AtHPT2    : ------------MELSISQSPRVRFSSLAPRFLAASHHHRPSVHLAGKFISLPRDVRF :  46
GmHPT2    : -----------MELSLSPTSHRVPSTIPTLNSAKLSSTKATKSQQPLFLGFSKHFNSIGL :  49
OsHPT2    : ---------------------MASLASPPLPCRAAATASRSGRPAPRLLGPPPPPASPLL :  39

                                                          Motif V
                    *         80         *        100         *        120
NoHPT     : ----------------------MSQSSQNSPLPRKPVQSYFHWLYAFWKFSRPHTIIGT :  37
AnHPT     : ----------------------MNQSSQDRPLRPKPLQSSFQWLYAFWKFSRPHTIIGT :  37
SyHPT     : ------------------------------------MATIQAFWRFSRPHTIIGT :  19
ZmHPT1    : EWKSSYCRISHRSLNTSVNASGQQLQSEPETHDSTTIWRAISSSLDAFYRFSRPHTVIGT : 118
TaHPT1    : DWKYSNHRISHQSINTSAKAG-QSLQPETEAHDPASFWKPISSSLDAFYRFSRPHTIIGT : 119
ApHPT1    : VKPRFTTCSRSQKLGHVKATSEHSLESGSEGYTPRSIWEAVLASLNVLYKFSRPHTIIGT : 114
GmHPT1-1  : SIEGGCTCKKCNIKFVVKATSEKSFESEPQAFDPKSILDSVKNSLDAFYRFSRPHTVIGT : 114
CpHPT1    : TRQRFTFHQNGHRTYLVKAVSGQSLESEPESYP-NNRWDYVKSAADAFYRFSRPHTIIGT : 112
AtHPT1    : VRPDGQGSSLLLYPKHKSRFRVNATAGQPEAFDSNSKQKSFRDSLDAFYRFSRPHTVIGT : 112
AtHPT2    : TSLSTSRMRSKFVSTNYRKISIRACSQVGAAESDDPVLDRIARFQNACWRFLRPHTIRGT : 106
GmHPT2    : HHHSYRCCSNAVPERPQRPSSIRACTGVGASGSDRPLAERLLDLKDACWRFLRPHTIRGT : 109
OsHPT2    : SSASARFPRAPCNAARWSRRDAVRVCSQAGAAGPAPLSKTLSDLKDSCWRFLRPHTIRGT :  99

                                                  Motif VI
                    *        140         *        160         *        180
NoHPT     : SLSVLSLYLIAIAISNNTASLFTTPGSLSPLFGAWIACLCGNVYIVGLNQLEDVDIDKIN :  97
AnHPT     : SLSVLGLYLISIAVSSTGFALTQIN----SVLGAWIACLCGNVYIVGLNQLEDIEIDKVN :  93
SyHPT     : TLSVWAVYLLTILGDGNSVNSPASLD---LVFGAWIACLLGNVYIVGLNQLWDVDIDRIN :  76
ZmHPT1    : ALSIVSVSLLAVQSLSDISPLFLTG-----LLEAVVAALFMNIYIVGLNQLFDIEIDKVN : 173
TaHPT1    : ALSIVSVSLLAVESLSDISPLFLTG-----LLEAVVAALFMNIYIVGLNQLFDIEIDKVN : 174
ApHPT1    : AMGIMSVSLLVVESLSDISPLFFVG-----LLEAVVAALFMNVYIVGLNQLFDIEIDKVN : 169
GmHPT1-1  : ALSIISVSLLAVEKISDISPLFFTG-----VLEAVVAALFMNIYIVGLNQLSDVEIDKIN : 169
CpHPT1    : ALSIVSVSLLAVEKLPELNSMFFTG-----LLEVIIAALFMNIYIVGLNQLSDIDIDKVN : 167
AtHPT1    : VLSILSVSFLAVEKVSDISPLLFTG-----ILEAVVAALMMNIYIVGLNQLSDVEIDKVN : 167
AtHPT2    : ALGSTALVTRALIENTHLIKWSLVLK----ALSGLIALICGNGYIVGINQIYDIGIDKVN : 162
GmHPT2    : ALGSFALVARALIENTNLIKWSLFFK----AFCGLFALICGNGYIVGINQIYDISIDKVN : 165
OsHPT2    : ALGSIALVARALIENPQLINWWLVFK----AFYGLVALICGNGYIVGINQIYDIRIDKVN : 155

                    *        200         *        220         *        240
NoHPT     : KPHLPLASGEFSQQTGQLIVASTGILALVMAWLTG---PFLFGMVTISLAIGTAYSLPPI : 154
AnHPT     : KPHLPLASGEFSRKQGRIIVILTGITAIVLAWLNG---PYLFGMVAVSLAIGTAYSLPPI : 150
SyHPT     : KPNLPLANGDFSIAQGRWIVGLCGVASLAIAWGLG---LWLGLTVGISLIIGTAYSVPPV : 133
ZmHPT1    : KPTLPLASGEYTLATGVAIVSVFAAMSFGLGWAVGSQPLFWALFISFVLGTAYSINLPYL : 233
TaHPT1    : KPTLPLASGEYSPATGVAIVSVFAAMSFGLGWVVGSPPLFWALFISFVLGTAYSVNLPYF : 234
ApHPT1    : KPDLPLASGEYSPRAGTAIVIASAIMSFGIGWLVGSWPLFWALFISFVLGTAYSINLPFL : 229
GmHPT1-1  : KPYLPLASGEYSFETGVTIVASFSILSFWLGWVVGSWPLFWALFVSFVLGTAYSINVPLL : 229
CpHPT1    : KPYLPLASGEFSVGTGVTIVTSFLIMSFWLGWVVGSWPLFWALFISFVLGTAYSIDMPML : 227
AtHPT1    : KPYLPLASGEYSVNTGIAIVASFSIMSFWLGWIVGSWPLFWALFVSFMLGTAYSINLPLL : 227
AtHPT2    : KPYLPIAAGDLSVQSAWLLVIFFAIAGLLVVGFNFG--PFITSLYSLGLFLGTIYSVPPL : 220
GmHPT2    : KPYLPIAAGDLSVQSAWFLVIFFAAAGLSIAGLNFG--PFIFSLYTLGLFLGTIYSVPPL : 223
OsHPT2    : KPYLPIAAGDLSVQTAWLLVVLFAAAGFSIVVTNFI--LFITSLYCLALFLGTIYSVPPF : 213
```

Fig. 24a

```
                                  *       260       *       280       *       300
NoHPT     : RLKQFPFWAALCIFSVRGTIVNLGLYLHYS-WALKQSQTIPPVVWVLTLFILVFTFAIAI : 213
AnHPT     : RLKQFPFWAALCIFSVRGTIVNLGLYLHFS-WLLQNKQSIPLEVWILTVFILIFTFAIAI : 209
SyHPT     : RLKRFSLLAALCILTVRGIVVNLGLFLFFR-IGLGYPPTLITDIWVLTLFILVFTVAIAI : 192
ZmHPT1    : RWKRFAVVAALCILAVRAVIVQLAFFLHIQTFVFRRPAVFSRDLLFATGFMTFFSVVIAL : 293
TaHPT1    : RWKRSAVVAALCILAVRAVIVQLAFFLHIQTFVFRRPAVFSKDLIFATAFMTFFSVVIAL : 294
ApHPT1    : RWKRSAVVAAICILAVRAVIVQLAFFLHIQSFVFKRPASFTRELIFATAFMSFFSVVIAL : 289
GmHPT1-1  : RWKRFAVLAAMCILAVRAVIVQLAFFLHIQTHVYKRPPVFSRSLIFATAFMSFFSVVIAL : 289
CpHPT1    : RWKRSAVVAALCILAVRAVIVQLAFFLHMQMHVYGRAAALSRPVIFATGFMSFFSIVIAL : 287
AtHPT1    : RWKRFALVAAMCILAVRAIIVQLAFYLHIQTHVFGRPILFTRDLIFATAFMSFFSVVIAL : 287
AtHPT2    : RMKRFPVAAFLIIATVRGFLLNFGVYHATR-AALGLPFQWSAPVAFITSFVTLFALVIAI : 279
GmHPT2    : RMKRFPVAAFLIIATVRGFLLNFGVYYATR-ASLGLAFEWSSPVVFITTFVTFFALVIAI : 282
OsHPT2    : RLKRYRAPACLIIATVRGFLRNLGVYYATR-AALGLTFQWSSPVAFITCFVTLFALVIAI : 272

                Motif VII
                                  *       320       *       340       *       360
NoHPT     : FKDIPDIIGDRLYNITTFTIKLGSQAVFNLALWVITVCYLGIILVGVLRIASVNPIFLIT : 273
AnHPT     : FKDIPDMIGDRLYNITTLTIQLGPQAVFNLAMWVLTVCYLGMVIIGVLRLGTINSVFLVV : 269
SyHPT     : FKDVPDMIGDRQFKIQTLTLQIGKQNVFRGTLILLTGCYLAMAIWGLWAAMPLNTAFLIV : 252
ZmHPT1    : FKDIPDIIGDRIFGIRSFSVRLGQKKVFWICVGLLEMAYSVAILMGATSSCLWSKTATIA : 353
TaHPT1    : FKDIPDIIGDRIFGIQSFSVRLGQSKVFWTCVGLLEVAYGVAILMGVTSSSLWSKSLTVV : 354
ApHPT1    : FKDIPDIDGDKIFGIHSFSVRLGQERVFWICIYLLEMAYTVVMVVGATSSCLWSKCLTVI : 349
GmHPT1-1  : FKDIPDIIGDKVFGIQSFSVRLGQKPVFWTCVILLEIAYGVAILVGAASPCLWSKIVTGL : 349
CpHPT1    : FKDIPDIIGDKIFGIRSFTVRLGQERVFWICISLLEMAYAVAILVGSTSPYLWSKVITVS : 347
AtHPT1    : FKDIPDIIGDKIFGIRSFSVTLGQKRVFWTCVTLLQMAYAVAILVGATSPFIWSKVISVV : 347
AtHPT2    : TKDLPDVIGDRKFQISTLATKLGVRNIAFLGSGLLLVNYVSAISLAFYMPQVFRGSLMIP : 339
GmHPT2    : TKDLPDVIGDRKYQISTFATKLGVRNIAFLGSGILLVNYIVSVILAAIYMPQAFRRWLLIP : 342
OsHPT2    : TKDLPDVIGDRKYQISTLATKLGVRNIAFLGSGLLIANYVAAIAVAFLMPQAFRRTVMVP : 332

                                                Motif VIII
                                  *       380       *       400       *
NoHPT     : AHLAILVWMWWRSLAVDLQ--DKSAIAQFYQFIWKLFFIEYLIFPIACFLA------- : 322
AnHPT     : THLVILCWMWMQSLAVDIH--DKTAIAQFYQFIWKLFFLEYLMFPIACLLA------- : 318
SyHPT     : SHLCLLALLWWRSRDVHLE--SKTEIASFYQFIWKLFFLEYLLYPLALWLPNFSNTIF : 308
ZmHPT1    : GHSILAAILWSCARSVDLT--SKAAITSFYMFIWKLFYAEYLLIPLVR---------- : 399
TaHPT1    : GHAILASILWSSARSIDLT--SKAAITSFYMLIWRLFYAEYLLIPLVR---------- : 400
ApHPT1    : GHAILGSLLWNRARSHGPM--TKTTITSFYMFVWKLFYAEYLLIPFVR---------- : 395
GmHPT1-1  : GHAVLASILWFHAKSVDLK--SKASITSFYMFIWKLFYAEYCLIPFVR---------- : 395
CpHPT1    : GHVVLASILWGRAKSIDFK--SKAALTSFYMFIWKLFYAEYLLIPLVR---------- : 393
AtHPT1    : GHVILATTLWARAKSVDLS--SKTEITSCYMFIWKLFYAEYLLLPFLK---------- : 393
AtHPT2    : AHVILASGLIFQTWVLEKANYTKEAISGYYRFIWNLFYAEYLLFPFL----------- : 386
GmHPT2    : AHTIFAISLIYQARILEQANYTKDAISGFYRFIWNLFYAEYALFPFI----------- : 389
OsHPT2    : VHAAILAVGIIFQTWVLEQAKYTKDAISQYYRFIWNLFYAEYTFFPLI----------- : 379
```

NoHPT : (SEQ ID NO: 1)
AnHPT : (SEQ ID NO: 2)
SyHPT : (SEQ ID NO: 3)
ZmHPT1 : (SEQ ID NO: 4)
TaHPT1 : (SEQ ID NO: 10)
ApHPT1 : (SEQ ID NO: 9)
GmHPT1-1 : (SEQ ID NO: 6)
CpHPT1 : (SEQ ID NO: 11)
AtHPT1 : (SEQ ID NO: 7)
AtHPT2 : (SEQ ID NO: 57)
GmHPT2 : (SEQ ID NO: 91)
OsHPT2 : (SEQ ID NO: 58)

Fig. 24b

Motif V (SEQ ID NO: 46)

```
NoHPT     : WKFSRPHTIIGT : 12
AnHPT     : WKFSRPHTIIGT : 12
SyHPT     : WRFSRPHTIIGT : 12
ZmHPT1    : YRFSRPHTVIGT : 12
TaHPT1    : YRFSRPHTIIGT : 12
ApHPT1    : YKFSRPHTIIGT : 12
GmHPT1-1  : YRFSRPHTVIGT : 12
CpHPT1    : YRFSRPHTIIGT : 12
AtHPT1    : YRFSRPHTVIGT : 12
AtHPT2    : WRFLRPHTIRGT : 12
GmHPT2    : WRFLRPHTIRGT : 12
OsHPT2    : WRFLRPHTIRGT : 12
```

Motif VI (SEQ ID NO: 47)

```
NoHPT     : NVYIVGLNQLEDVDIDKINKPHLPLA : 26
AnHPT     : NVYIVGLNQLEDIEIDKVNKPHLPLA : 26
SyHPT     : NVYIVGLNQLWDVDIDRINKPNLPLA : 26
ZmHPT1    : NIYIVGLNQLFDIEIDKVNKPTLPLA : 26
TaHPT1    : NIYIVGLNQLFDIEIDKVNKPTLPLA : 26
ApHPT1    : NVYIVGLNQLFDIEIDKVNKPDLPLA : 26
GmHPT1-1  : NIYIVGLNQLSDVEIDKINKPYLPLA : 26
CpHPT1    : NIYIVGLNQLSDIDIDKVNKPYLPLA : 26
AtHPT1    : NIYIVGLNQLSDVEIDKVNKPYLPLA : 26
AtHPT2    : NGYIVGINQIYDIGIDKVNKPYLPIA : 26
GmHPT2    : NGYIVGINQIYDISIDKVNKPYLPIA : 26
OsHPT2    : NGYIVGINQIYDIRIDKVNKPYLPIA : 26
```

*

Motif VII (SEQ ID NO: 48)

```
NoHPT     : IAIFKDIPDIIGDR : 14
AnHPT     : IAIFKDIPDMIGDR : 14
SyHPT     : IAIFKDVPDMIGDR : 14
ZmHPT1    : IALFKDIPDIIGDR : 14
TaHPT1    : IALFKDIPDIIGDR : 14
ApHPT1    : IALFKDIPDIDGDK : 14
GmHPT1-1  : IALFKDIPDIIGDK : 14
CpHPT1    : IALFKDIPDIIGDK : 14
AtHPT1    : IALFKDIPDIIGDK : 14
AtHPT2    : IAITKDLPDVIGDR : 14
GmHPT2    : IAITKDLPDVIGDR : 14
OsHPT2    : IAITKDLPDVIGDR : 14
```

Fig.25a

Motif VIII (SEQ ID NO: 49)

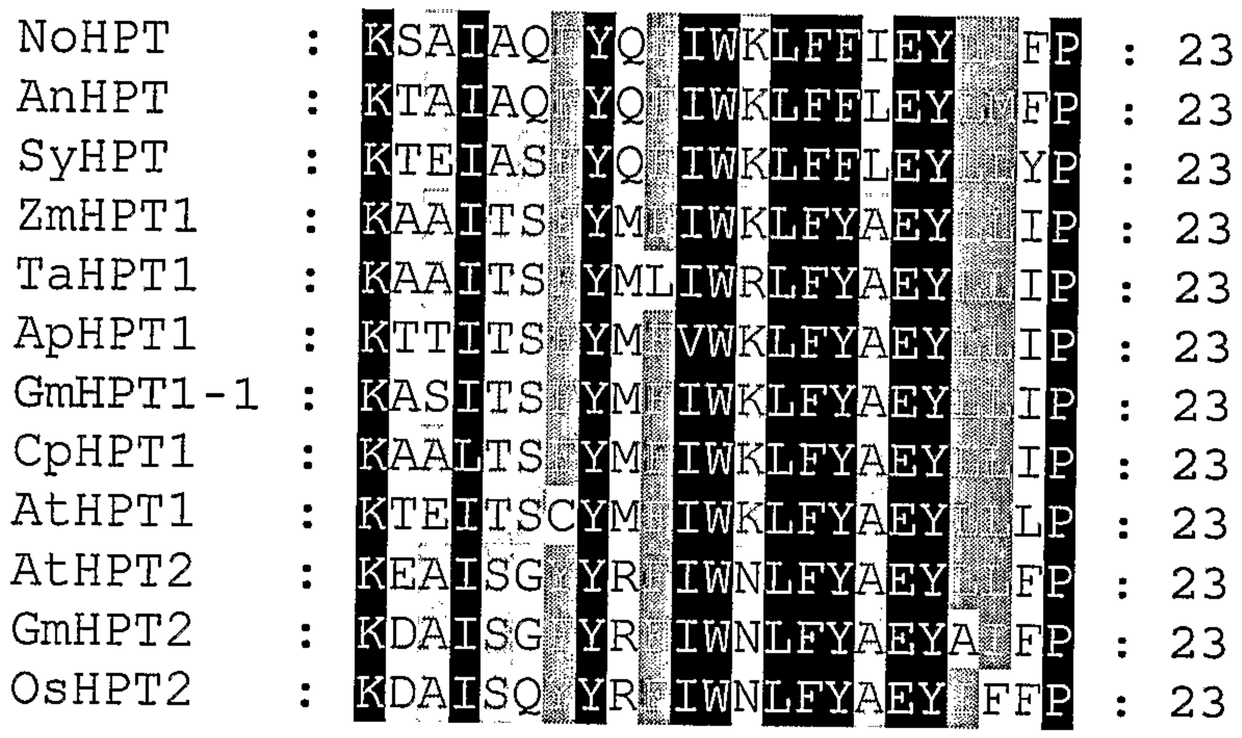


Fig. 25b

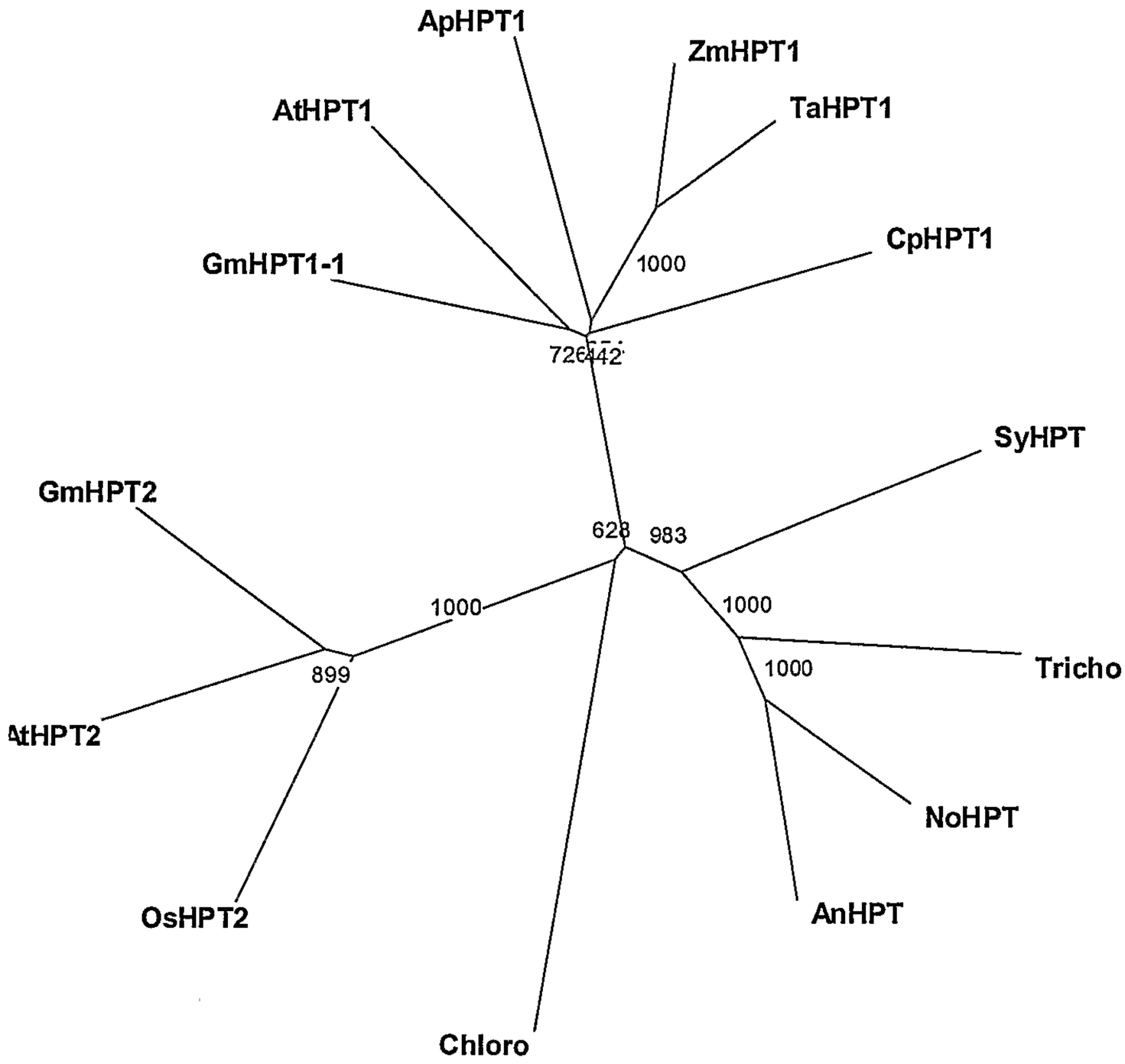

ApHPT1
ZmHPT1
AtHPT1
TaHPT1
CpHPT1
GmHPT1-1
1000
726442
SyHPT
GmHPT2
628
983
1000
1000
1000
899
Tricho
AtHPT2
NoHPT
OsHPT2
AnHPT
Chloro
0.1

Fig. 26

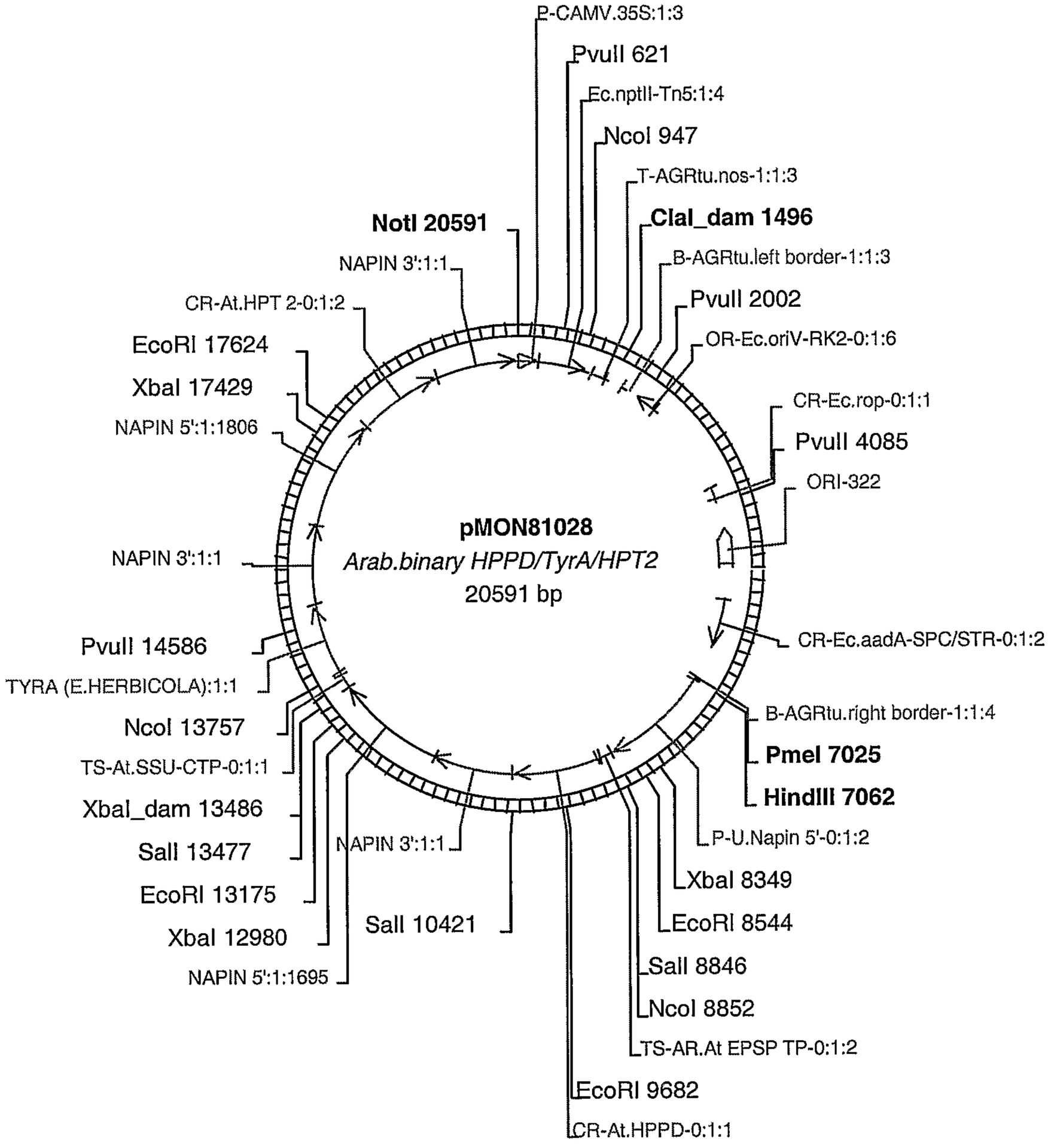

P-CAMV.35S:1:3
PvuII 621
Ec.nptII-Tn5:1:4
NcoI 947
T-AGRtu.nos-1:1:3
NotI 20591
ClaI_dam 1496
NAPIN 3':1:1
B-AGRtu.left border-1:1:3
CR-At.HPT 2-0:1:2
PvuII 2002
EcoRI 17624
OR-Ec.oriV-RK2-0:1:6
XbaI 17429
CR-Ec.rop-0:1:1
NAPIN 5':1:1806
PvuII 4085
ORI-322
pMON81028
NAPIN 3':1:1
Arab.binary HPPD/TyrA/HPT2
20591 bp
PvuII 14586
CR-Ec.aadA-SPC/STR-0:1:2
TYRA (E.HERBICOLA):1:1
B-AGRtu.right border-1:1:4
NcoI 13757
PmeI 7025
TS-At.SSU-CTP-0:1:1
HindIII 7062
XbaI_dam 13486
NAPIN 3':1:1
P-U.Napin 5'-0:1:2
SalI 13477
XbaI 8349
EcoRI 13175
EcoRI 8544
SalI 10421
XbaI 12980
SalI 8846
NcoI 8852
NAPIN 5':1:1695
TS-AR.At EPSP TP-0:1:2
EcoRI 9682
CR-At.HPPD-0:1:1

Fig. 27

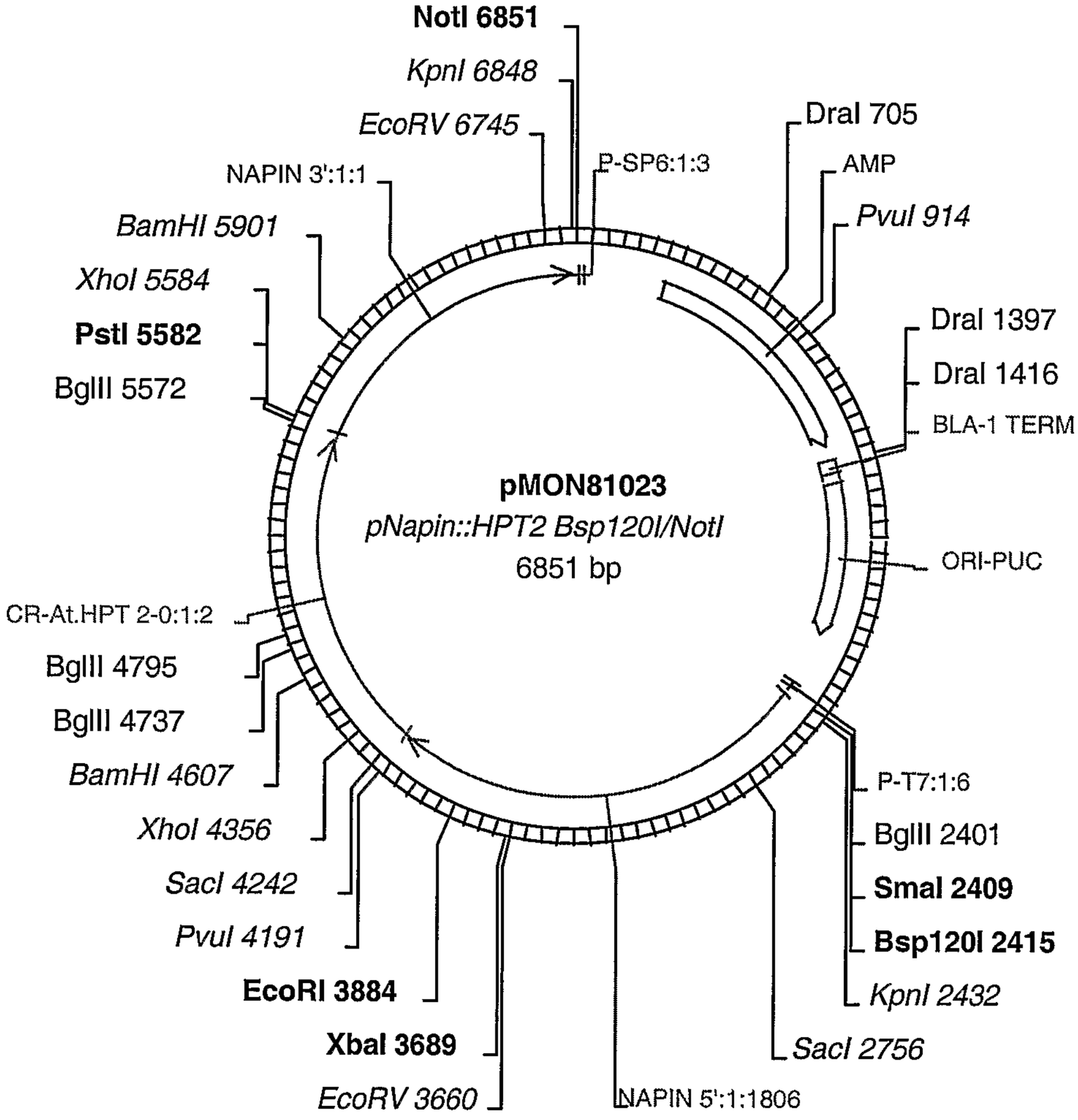

NotI 6851
KpnI 6848
EcoRV 6745
NAPIN 3':1:1
BamHI 5901
XhoI 5584
PstI 5582
BglII 5572
P-SP6:1:3
DraI 705
AMP
PvuI 914
DraI 1397
DraI 1416
BLA-1 TERM
pMON81023
pNapin::HPT2 Bsp120I/NotI
6851 bp
ORI-PUC
CR-At.HPT 2-0:1:2
BglII 4795
BglII 4737
BamHI 4607
XhoI 4356
SacI 4242
PvuI 4191
EcoRI 3884
XbaI 3689
EcoRV 3660
NAPIN 5':1:1806
SacI 2756
P-T7:1:6
BglII 2401
SmaI 2409
Bsp120I 2415
KpnI 2432

Fig. 28

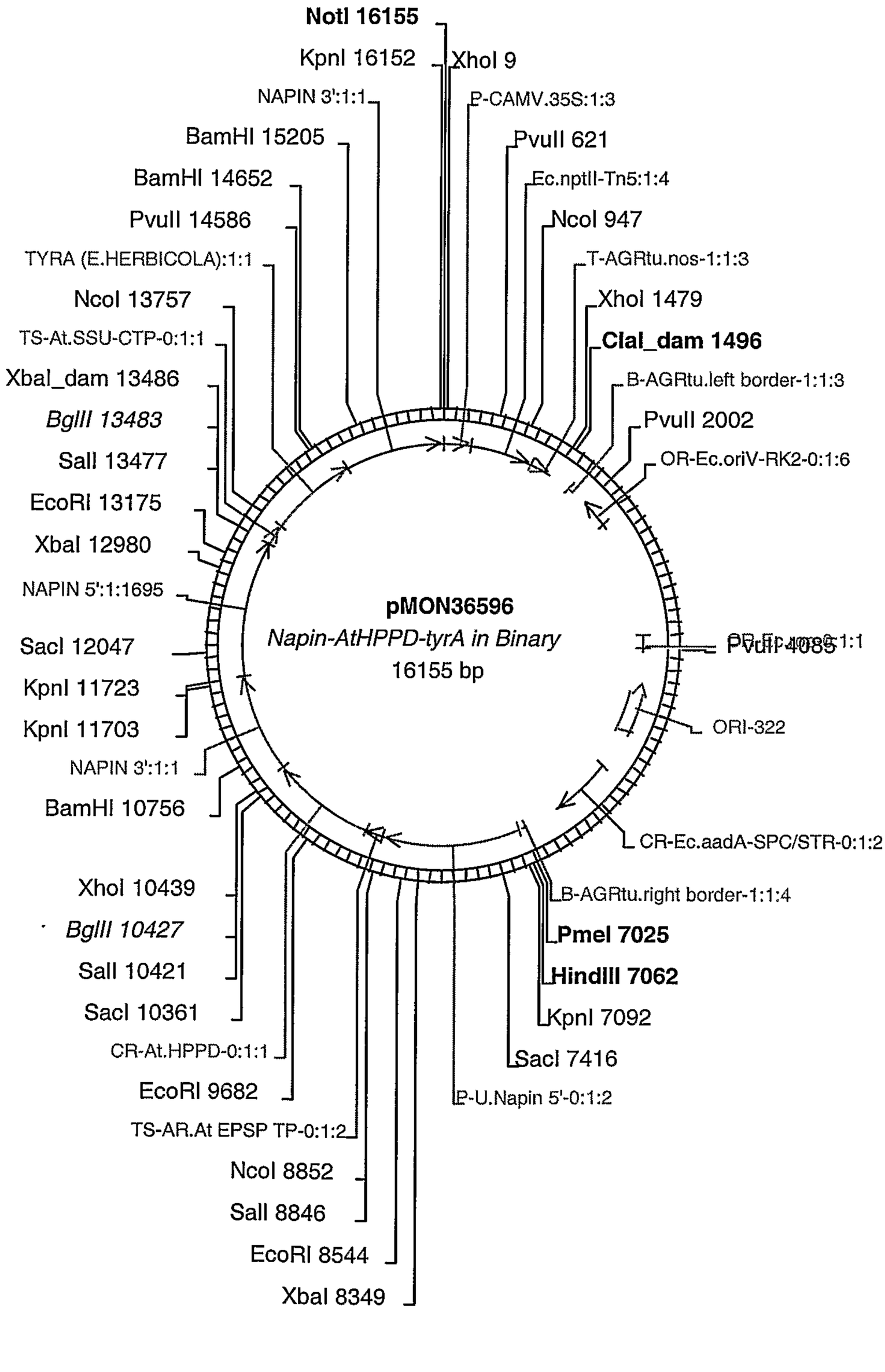

NotI 16155
KpnI 16152
XhoI 9
NAPIN 3':1:1
P-CAMV.35S:1:3
BamHI 15205
PvuII 621
BamHI 14652
Ec.nptII-Tn5:1:4
PvuII 14586
NcoI 947
TYRA (E.HERBICOLA):1:1
T-AGRtu.nos-1:1:3
NcoI 13757
XhoI 1479
TS-At.SSU-CTP-0:1:1
ClaI_dam 1496
XbaI_dam 13486
B-AGRtu.left border-1:1:3
BglII 13483
PvuII 2002
SalI 13477
OR-Ec.oriV-RK2-0:1:6
EcoRI 13175
XbaI 12980
NAPIN 5':1:1695
pMON36596
Napin-AtHPPD-tyrA in Binary
16155 bp
SacI 12047
PvuII 4085
KpnI 11723
KpnI 11703
ORI-322
NAPIN 3':1:1
BamHI 10756
CR-Ec.aadA-SPC/STR-0:1:2
XhoI 10439
B-AGRtu.right border-1:1:4
BglII 10427
PmeI 7025
SalI 10421
HindIII 7062
SacI 10361
KpnI 7092
CR-At.HPPD-0:1:1
SacI 7416
EcoRI 9682
P-U.Napin 5'-0:1:2
TS-AR.At EPSP TP-0:1:2
NcoI 8852
SalI 8846
EcoRI 8544
XbaI 8349

Fig. 29

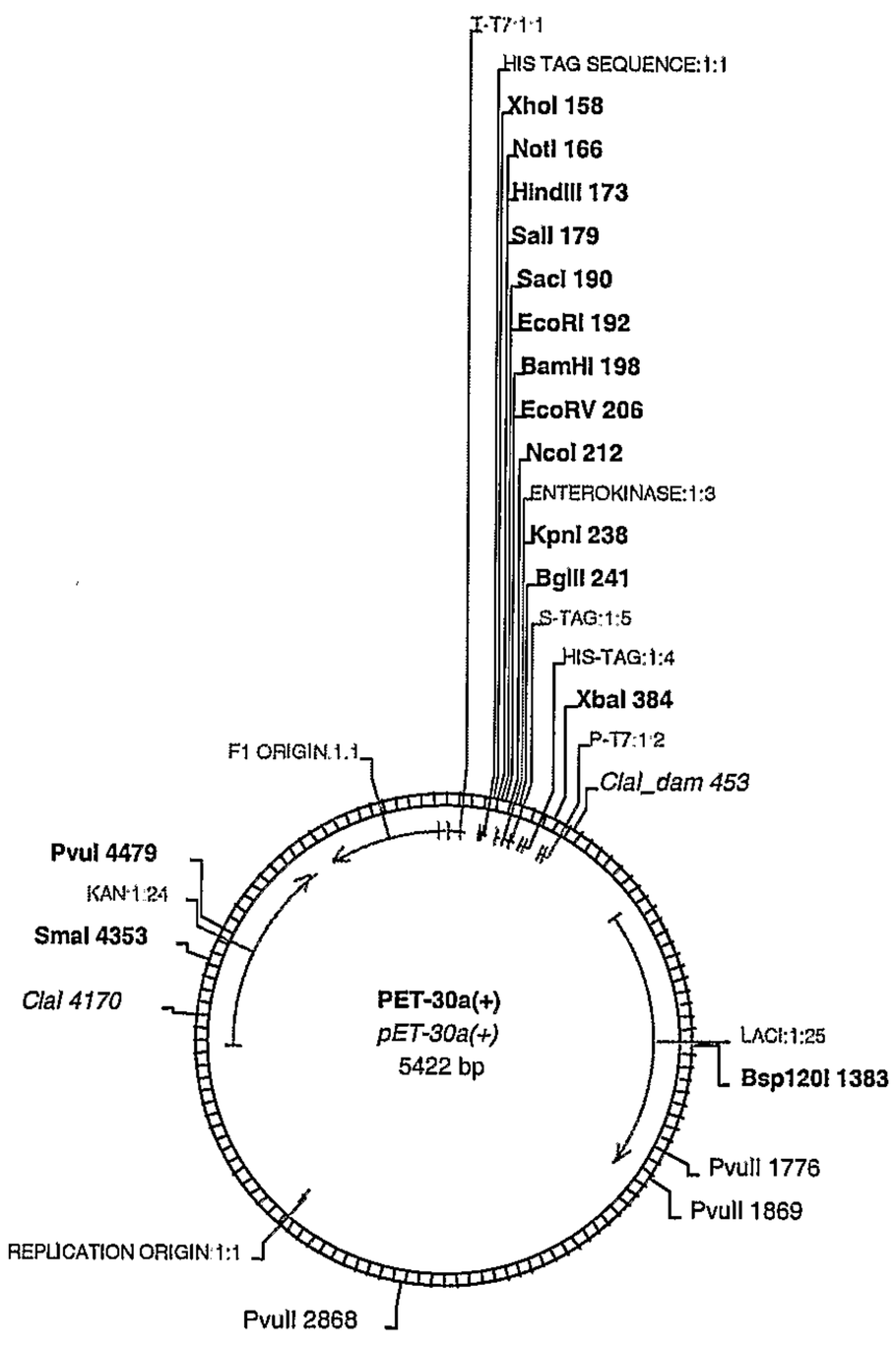

T7-1:1
HIS TAG SEQUENCE:1:1
XhoI 158
NotI 166
HindIII 173
SalI 179
SacI 190
EcoRI 192
BamHI 198
EcoRV 206
NcoI 212
ENTEROKINASE:1:3
KpnI 238
BglII 241
S-TAG:1:5
HIS-TAG:1:4
XbaI 384
P-T7:1:2
ClaI_dam 453
F1 ORIGIN:1:1
PvuI 4479
KAN:1:24
SmaI 4353
ClaI 4170
PET-30a(+)
pET-30a(+)
5422 bp
LACI:1:25
Bsp120I 1383
PvuII 1776
PvuII 1869
REPLICATION ORIGIN:1:1
PvuII 2868

Fig. 30

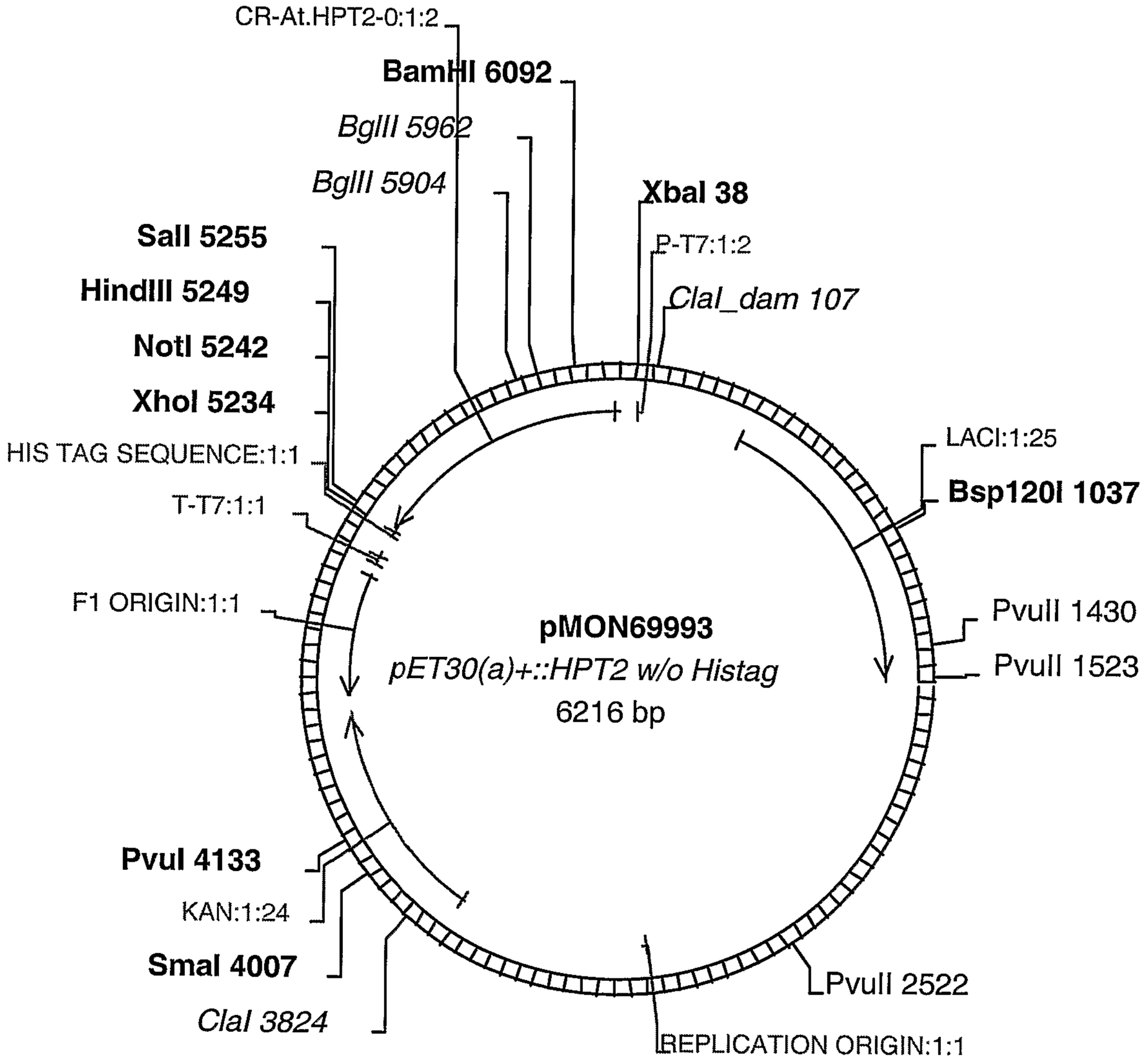

CR-At.HPT2-0:1:2
BamHI 6092
BglII 5962
BglII 5904
XbaI 38
P-T7:1:2
ClaI_dam 107
SalI 5255
HindIII 5249
NotI 5242
XhoI 5234
HIS TAG SEQUENCE:1:1
T-T7:1:1
F1 ORIGIN:1:1
LACI:1:25
Bsp120I 1037
PvuII 1430
PvuII 1523
pMON69993
pET30(a)+::HPT2 w/o Histag
6216 bp
PvuI 4133
KAN:1:24
SmaI 4007
ClaI 3824
PvuII 2522
REPLICATION ORIGIN:1:1

Fig. 31

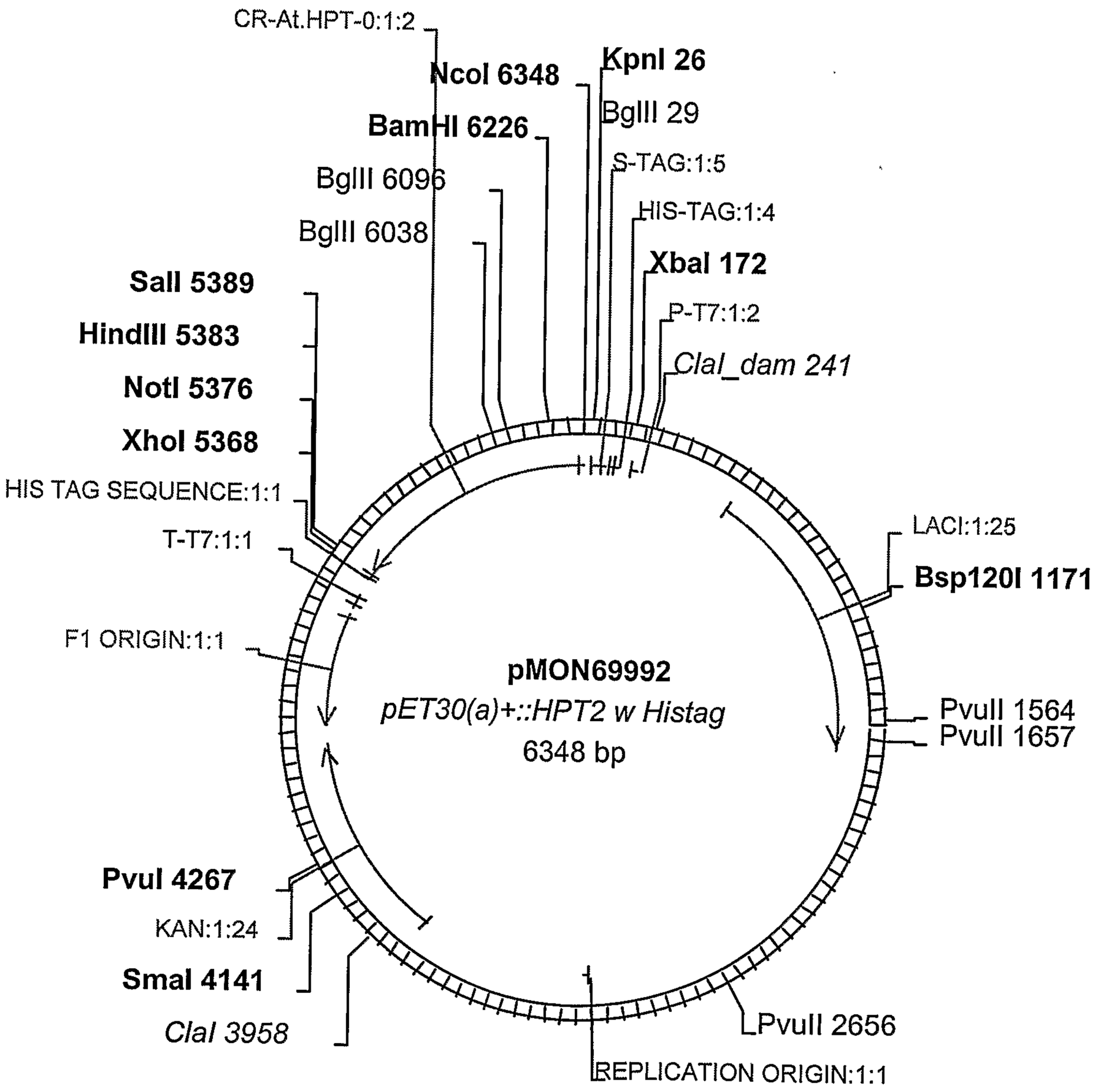

CR-At.HPT-0:1:2
NcoI 6348
KpnI 26
BamHI 6226
BglII 29
BglII 6096
S-TAG:1:5
BglII 6038
HIS-TAG:1:4
SalI 5389
XbaI 172
HindIII 5383
P-T7:1:2
NotI 5376
ClaI_dam 241
XhoI 5368
HIS TAG SEQUENCE:1:1
LACI:1:25
T-T7:1:1
Bsp120I 1171
F1 ORIGIN:1:1
pMON69992
pET30(a)+::HPT2 w Histag
6348 bp
PvuII 1564
PvuII 1657
PvuI 4267
KAN:1:24
SmaI 4141
ClaI 3958
PvuII 2656
REPLICATION ORIGIN:1:1

Fig. 32

```
                    *         20         *         40         *         60
NoHPT     : ------------------------------------------------------------ :   -
AnHPT     : ------------------------------------------------------------ :   -
Tricho    : ------------------------------------------------------------ :   -
SyHPT     : ------------------------------------------------------------ :   -
ZmHPT1    : --MDALRLRPSLLPVRPGAARPRDHFLPPCCSIQRNGEGRICFSSQRTQGPTLHHHQKFF : 58
TaHPT1    : MDSLRLRPSSLRSAPGAAAARRRDHILPSFCSIQRNGKGRVTLSIQASKGPTINHCKKFL : 60
ApHPT1    : ------MLSMDSLLTKPVVIPLPSPVCSLPILRGSSAPGQYSCRNYNPIRIQRCLVNYEH : 54
GmHPT1-1  : ------MDSMLLRSFPNINNASSLATTGSYLPNASWHNRKIQKEYNFLRFRWPSLNHHYK : 54
CpHPT1    : -------MRMESLLLNSFSPSPAGGKICRADTYKKAYFATARCNTLNSLNKNTGEYHLSR : 53
AtHPT1    : --------MESLLSSSSLVSAAGGFCWKKQNLKLHSLSEIRVLRCDSSKVVAKPKFRNNL : 52
Chloro    : ------------------------------------------------------------ :   -
AtHPT2    : --------------MELSISQSPRVRFSSLAPRFLAASHHHRPSVHLAGKFISLPRDVRF : 46
GmHPT2    : ----------MELSLSPTSHRVPSTIPTLNSAKLSSTKATKSQQPLFLGFSKHFNSIGL : 49
OsHPT2    : --------------------MASLASPPLPCRAAATASRSGRPAPRLLGPPPPPASPLL : 39

                                                          Motif IX

                    *         80         *         100        *         120
NoHPT     : ----------------------MSQSSQNSPLPRKPVQSYFHWLYAFWKFSRPHTIWCT : 37
AnHPT     : ----------------------MNQSSQDRPLRPKPLQSSFQWLYAFWKFSRPHTIWCT : 37
Tricho    : --------------------MGKIAGSQQGKITTNWLQKYVPWLYSFWKFARPHTIWCT : 39
SyHPT     : ------------------------------------MATIQAFWKFSRPHTIWCT : 19
ZmHPT1    : EWKSSYCRISHRSLNTSVNASGQQLQSEPETHDSTTIWRAISSSLDAFWKFSRPHTVWCT : 118
TaHPT1    : DWKYSNHRISHQSINTSAKAG-QSLQPETEAHDPASFWKPISSSLDAFWKFSRPHTIWCT : 119
ApHPT1    : VKPRFTTCSRSQKLGHVKATSEHSLESGSEGYTPRSIWEAVLASLNVLWKFSRPHTIWCT : 114
GmHPT1-1  : SIEGGCTCKKCNIKFVVKATSEKSFESEPQAFDPKSILDSVKNSLDAFWKFSRPHTVWCT : 114
CpHPT1    : TRQRFTFHQNGHRTYLVKAVSGQSLESEPESYP-NNRWDYVKSAADAFWKFSRPHTIWCT : 112
AtHPT1    : VRPDGQGSSLLLYPKHKSRFRVNATAGQPEAFDSNSKQKSFRDSLDAFWKFSRPHTVWCT : 112
Chloro    : ----------------------------------------MRKQLRLLIEFARPHTVWAT : 20
AtHPT2    : TSLSTSRMRSKFVSTNYRKISIRACSQVGAAESDDPVLDRIARFQNACWRFLRPHTIRCT : 106
GmHPT2    : HHHSYRCCSNAVPERPQRPSSIRACTGVGASGSDRPLAERLLDLKDACWRFLRPHTIRCT : 109
OsHPT2    : SSASARFPRAPCNAARWSRRDAVRVCSQAGAAGPAPLSKTLSDLKDSCWRFLRPHTIRCT : 99
```

Fig. 33a

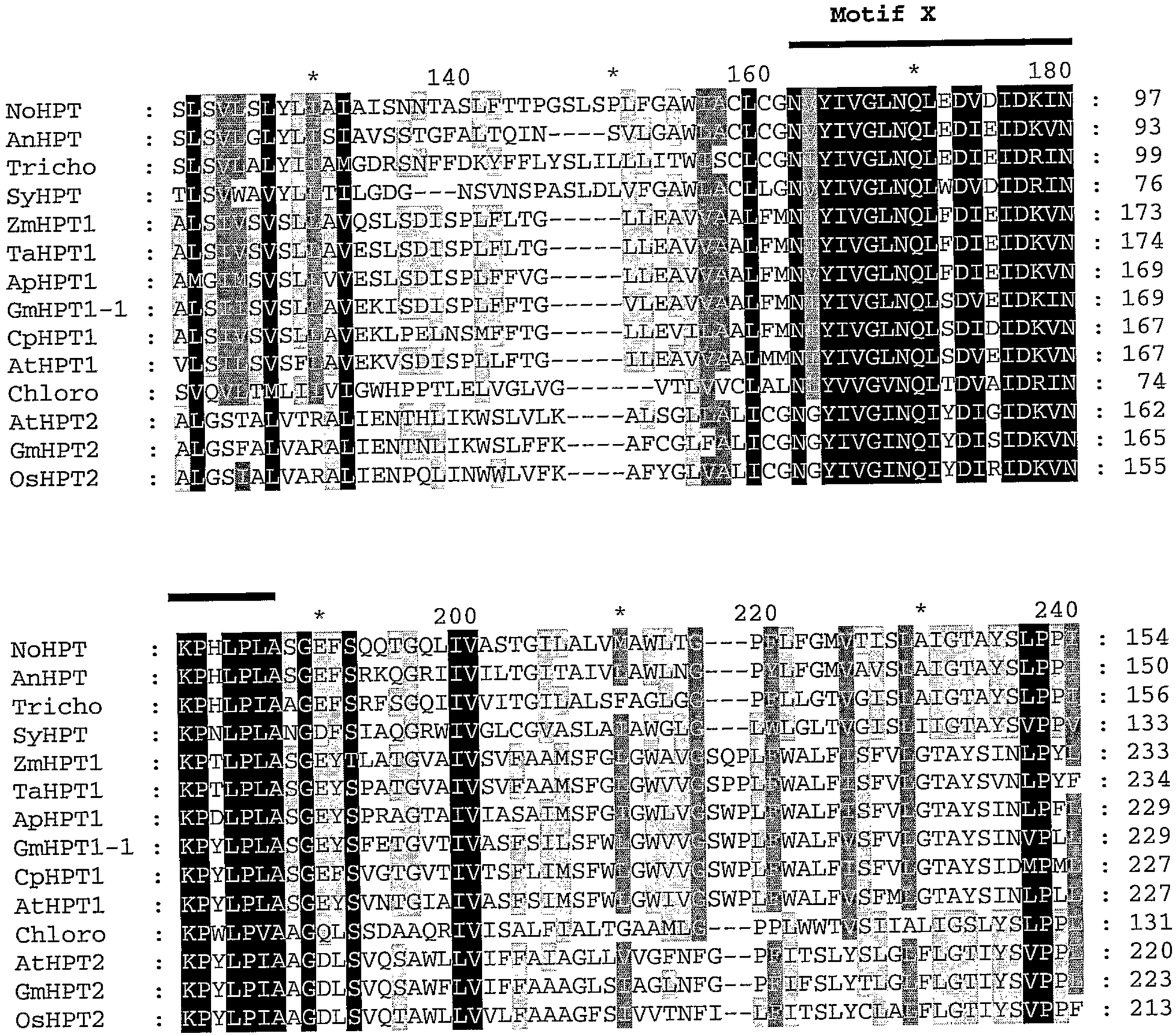

Motif X
NoHPT
AnHPT
Tricho
SyHPT
ZmHPT1
TaHPT1
ApHPT1
GmHPT1-1
CpHPT1
AtHPT1
Chloro
AtHPT2
GmHPT2
OsHPT2

Fig. 33b

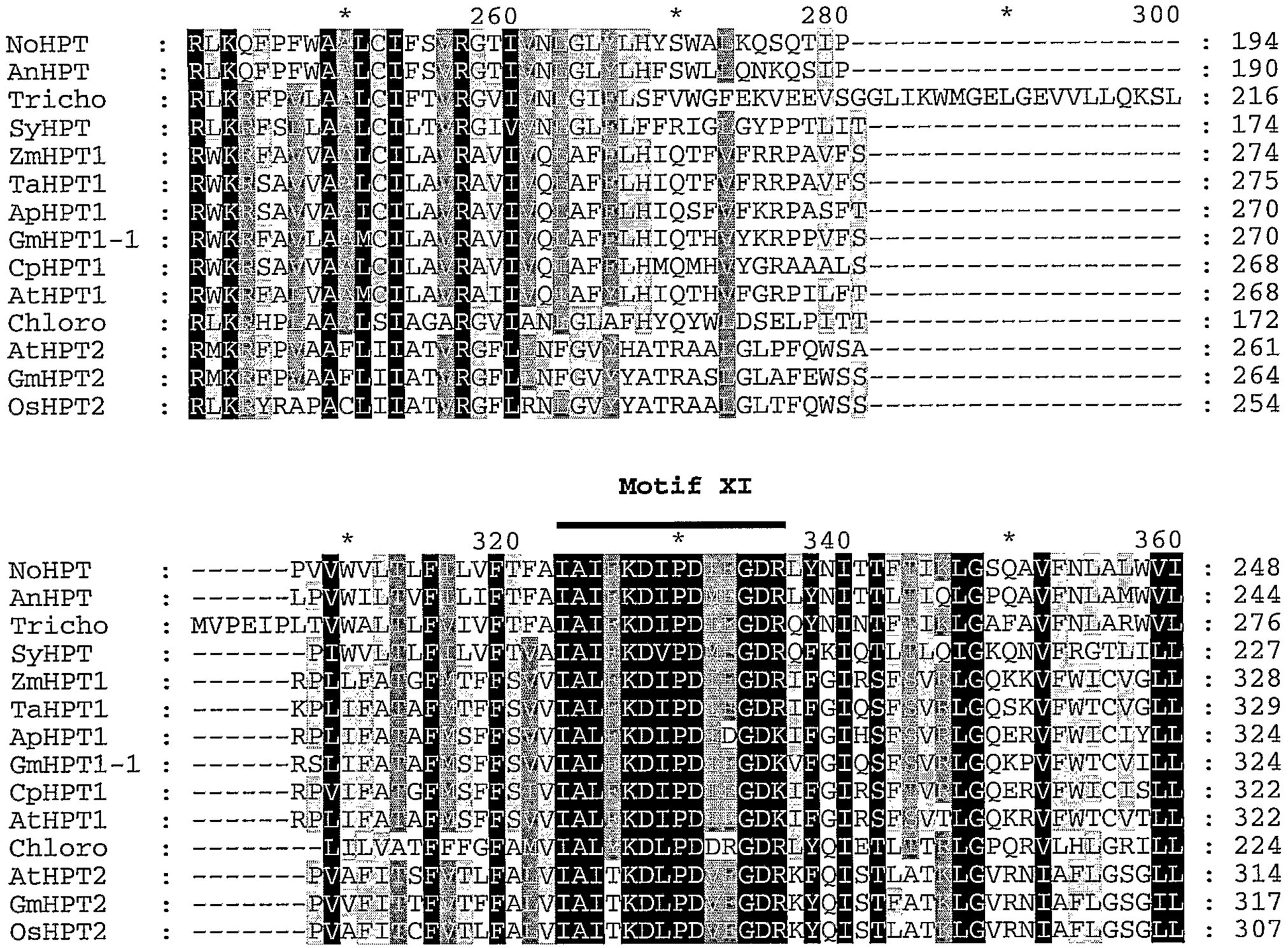

Motif XI
NoHPT : RLKQFPFWAALCIFSVRGTIVNLGLVLHYSWALKQSQTIP---------- : 194
AnHPT : RLKQFPFWAALCIFSVRGTIVNLGLVLHFSWLLQNKQSIP---------- : 190
Tricho : RLKRFPVLAALCIFTVRGVIVNLGIFLSFVWGFEKVEEVSGGLIKWMGELGEVVLLQKSL : 216
SyHPT : RLKRFSLLAALCILTVRGIVVNLGLFLFFRIGLGYPPTLIT---------- : 174
ZmHPT1 : RWKRFAVVAALCILAVRAVIVQLAFFLHIQTFVFRRPAVFS---------- : 274
TaHPT1 : RWKRSAVVAALCILAVRAVIVQLAFFLHIQTFVFRRPAVFS---------- : 275
ApHPT1 : RWKRSAVVAICILAVRAVIVQLAFFLHIQSFVFKRPASFT---------- : 270
GmHPT1-1 : RWKRFAVLAAMCILAVRAVIVQLAFFLHIQTHVYKRPPVFS---------- : 270
CpHPT1 : RWKRSAVVALCILAVRAVIVQLAFFLHMQMHVYGRAAALS---------- : 268
AtHPT1 : RWKRFALVAAMCILAVRAIIVQLAFYLHIQTHVFGRPILFT---------- : 268
Chloro : RLKRHPLAAALSIAGARGVIANLGLAFHYQYWLDSELPITT---------- : 172
AtHPT2 : RMKRFPVAAFLIIATVRGFLLNFGVVHATRAALGLPFQWSA---------- : 261
GmHPT2 : RMKRFPVAAFLIIATVRGFLLNFGVVYATRASLGLAFEWSS---------- : 264
OsHPT2 : RLKRYRAPACLIIATVRGFLRNLGVVYATRAALGLTFQWSS---------- : 254
NoHPT : ------PVVWVLILFILVFTFAIAIFKDIPDFRGDRLYNITTFFILLGSQAVFNLALWVI : 248
AnHPT : ------LPVWILIVFFLIFTFAIAIFKDIPDVRGDRLYNITTLIQLGPQAVFNLAMWVL : 244
Tricho : MVPEIPLTVWALILFVIVFTFAIAIFKDIPDFRGDRQYNINTFFILLGAFAVFNLARWVL : 276
SyHPT : -------PIWVLILFILVFTVAIAIFKDVPDVRGDRQFKIQTLILQIGKQNVFRGTLILL : 227
ZmHPT1 : ------RPLLFAIGFVTFFSVVIALFKDIPDVRGDRIFGIRSFSVRLGQKKVFWICVGLL : 328
TaHPT1 : ------KPLIFAIAFVTFFSVVIALFKDIPDVRGDRIFGIQSFSVRLGQSKVFWTCVGLL : 329
ApHPT1 : ------RPLIFAIAFVSFFSVVIALFKDIPDIDGDKIFGIHSFSVRLGQERVFWICIYLL : 324
GmHPT1-1 : ------RSLIFAIAFVSFFSVVIALFKDIPDVRGDKVFGIQSFSVRLGQKPVFWTCVILL : 324
CpHPT1 : ------RPVIFAIGFVSFFSIVIALFKDIPDVRGDKIFGIRSFTVRLGQERVFWICISLL : 322
AtHPT1 : ------RPLIFAIAFVSFFSVVIALFKDIPDVRGDKIFGIRSFSVTLGQKRVFWTCVTLL : 322
Chloro : --------LILVATFFFGFAMVIALYKDLPDDRGDRLYQIETLTTRLGPQRVLHLGRILL : 224
AtHPT2 : -------PVAFIISFVTLFALVIAITKDLPDVEGDRKFQISTLATKLGVRNIAFLGSGLL : 314
GmHPT2 : -------PVVFIITFVTFFALVIAITKDLPDVEGDRKYQISTFATKLGVRNIAFLGSGIL : 317
OsHPT2 : -------PVAFIICFVTLFALVIAITKDLPDVEGDRKYQISTLATILGVRNIAFLGSGLL : 307

Fig. 33c

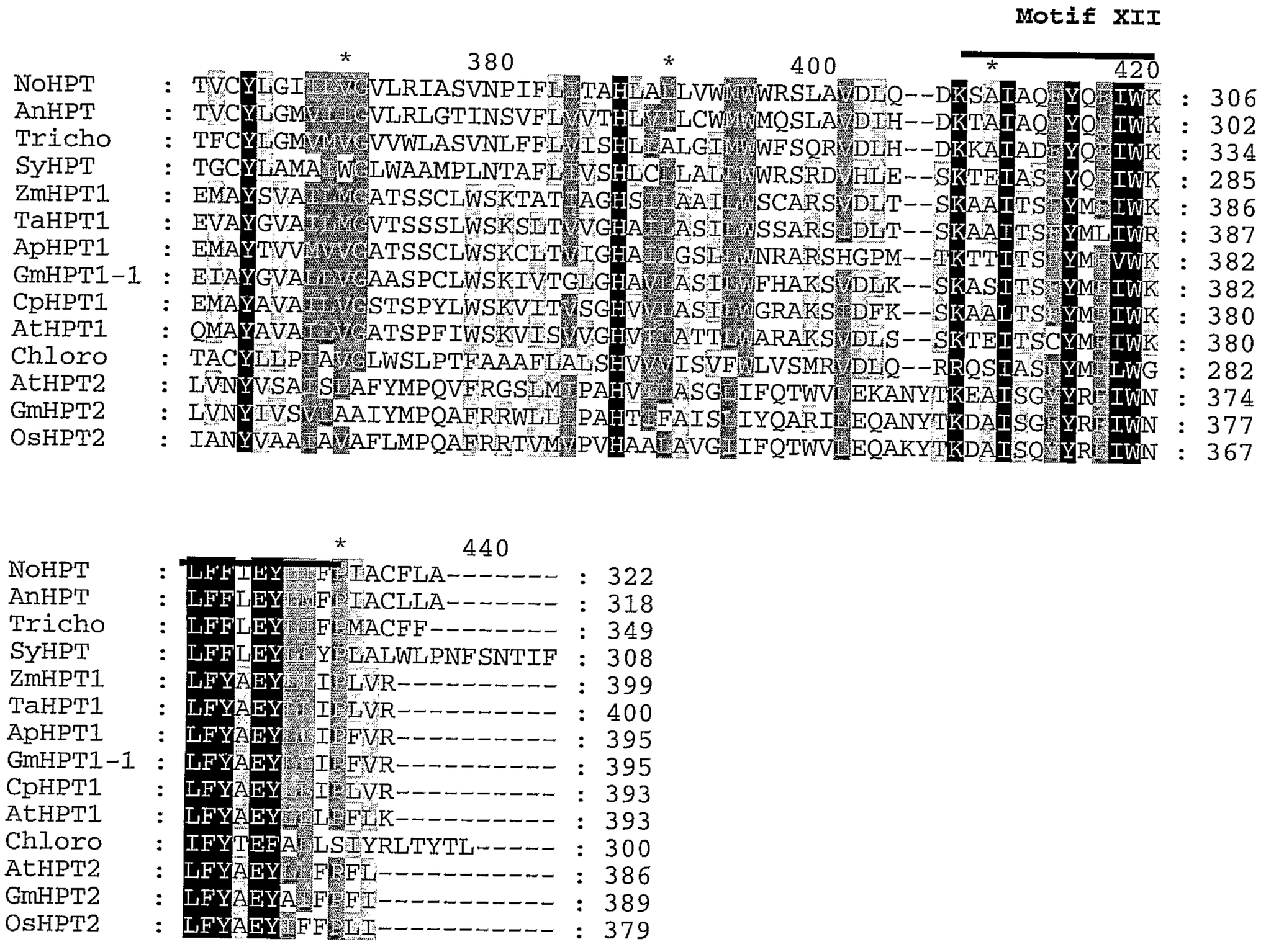


Fig. 33d

CA 02478957 2004-09-09

Motif IX (SEQ ID NO: 92)

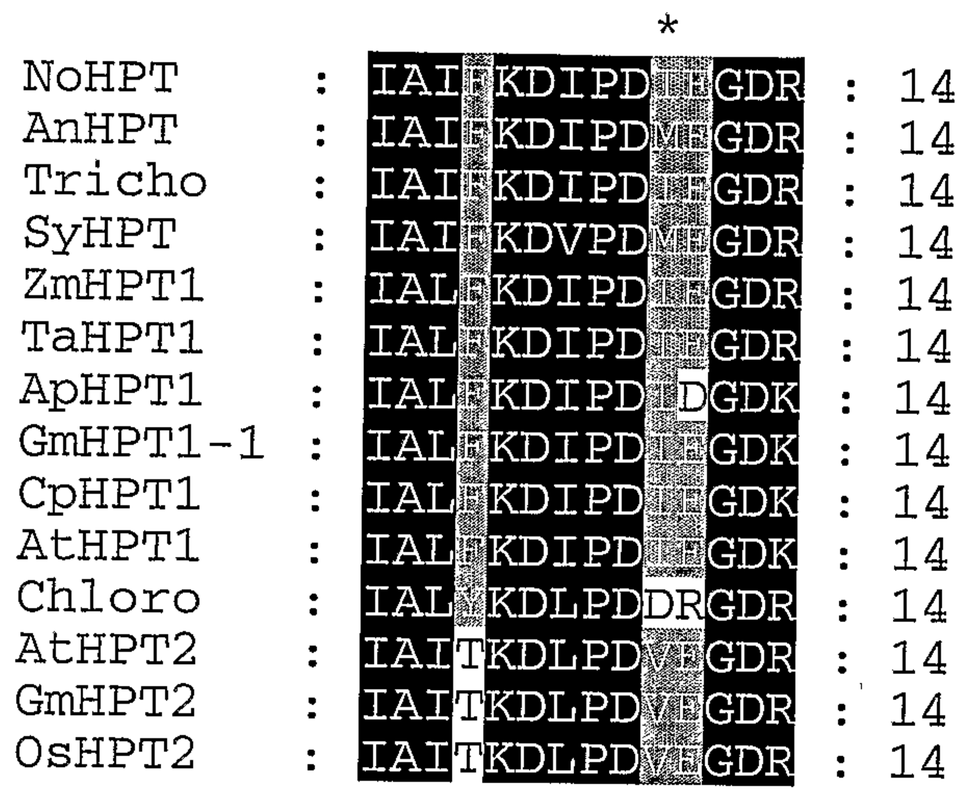
```
                          *
NoHPT     :  WKFSRPHTICT  : 12
AnHPT     :  WKFSRPHTICT  : 12
Tricho    :  WKFARPHTICT  : 12
SyHPT     :  WRFSRPHTICT  : 12
ZmHPT1    :  YRFSRPHTVCT  : 12
TaHPT1    :  YRFSRPHTICT  : 12
ApHPT1    :  YKFSRPHTICT  : 12
GmHPT1-1  :  YRFSRPHTVCT  : 12
CpHPT1    :  YRFSRPHTICT  : 12
AtHPT1    :  YRFSRPHTVCT  : 12
Chloro    :  IEFARPHTVIAT  : 12
AtHPT2    :  WRFLRPHTIRCT  : 12
GmHPT2    :  WRFLRPHTIRCT  : 12
OsHPT2    :  WRFLRPHTIRCT  : 12
```

Motif X (SEQ ID NO: 93)

```
                          *          20
NoHPT     :  NVYIVGLNQLEDVDIDKINKPHLPLA  : 26
AnHPT     :  NVYIVGLNQLEDIEIDKVNKPHLPLA  : 26
Tricho    :  NIYIVGLNQLEDIEIDRINKPHLPIA  : 26
SyHPT     :  NVYIVGLNQLWDVDIDRINKPNLPLA  : 26
ZmHPT1    :  NIYIVGLNQLFDIEIDKVNKPTLPLA  : 26
TaHPT1    :  NIYIVGLNQLFDIEIDKVNKPTLPLA  : 26
ApHPT1    :  NVYIVGLNQLFDIEIDKVNKPDLPLA  : 26
GmHPT1-1  :  NIYIVGLNQLSDVEIDKINKPYLPLA  : 26
CpHPT1    :  NIYIVGLNQLSDIDIDKVNKPYLPLA  : 26
AtHPT1    :  NIYIVGLNQLSDVEIDKVNKPYLPLA  : 26
Chloro    :  NLYVVGVNQLTDVAIDRINKPWLPVA  : 26
AtHPT2    :  NGYIVGINQIYDIGIDKVNKPYLPIA  : 26
GmHPT2    :  NGYIVGINQIYDISIDKVNKPYLPIA  : 26
OsHPT2    :  NGYIVGINQIYDIRIDKVNKPYLPIA  : 26
```

Fig. 34a

Motif XI (SEQ ID NO: 94)



Motif XII (SEQ ID NO: 95)

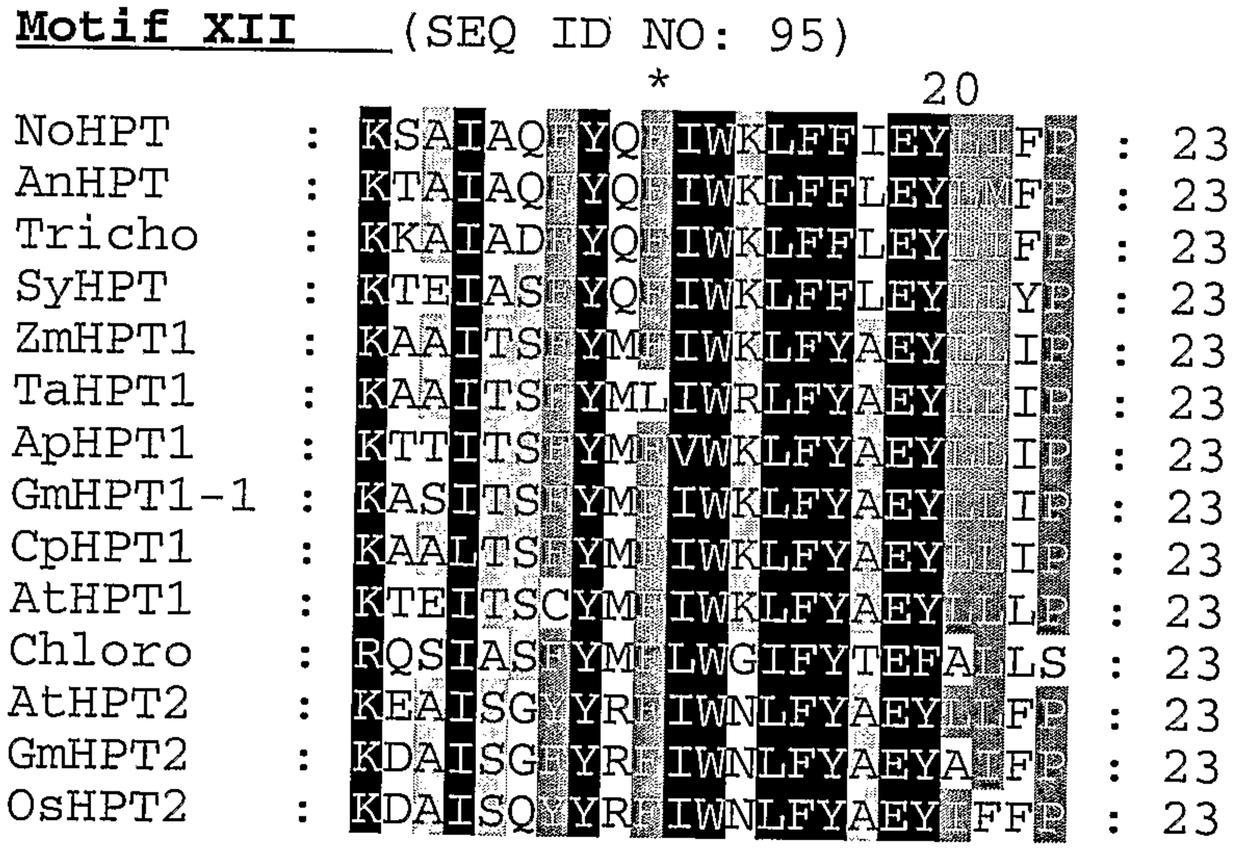


Fig. 34b

Motif I

```
                          *
Corn          :  AAYRFSRPHTVIGT  : 14
Wheat         :  AAYRFSRPHTIIGT  : 14
Leek          :  VLYKFSRPHTIIGT  : 14
soy-ppt2      :  AAYRFSRPHTVIGT  : 14
soy-ppt1      :  AAYRFSRPHTVIGT  : 14
Cuphea        :  AAYRFSRPHTIIGT  : 14
Arabidopsi    :  AAYRFSRPHTVIGT  : 14
Nostoc_ppt    :  AAWKFSRPHTIIGT  : 14
anabaena_p    :  AAWKFSRPHTIIGT  : 14
Synechocys    :  AAWRFSRPHTIIGT  : 14
                 af54FSRPHT6IGT
```

Motif II

```
                          *          20
Corn          :  NIYIVGLNQLFDIEIDKVNKPTLPLA  : 26
Wheat         :  NIYIVGLNQLFDIEIDKVNKPTLPLA  : 26
Leek          :  NVYIVGLNQLFDIEIDKVNKPDLPLA  : 26
soy-ppt2      :  NIYIVGLNQLSDVEIDKINKPYLPLA  : 26
soy-ppt1      :  NIYIVGLNQLSDVEIDKINKPYLPLA  : 26
Cuphea        :  NIYIVGLNQLSDIDIDKVNKPYLPLA  : 26
Arabidopsi    :  NIYIVGLNQLSDVEIDKVNKPYLPLA  : 26
Nostoc_ppt    :  NVYIVGLNQLEDVDIDKINKPHLPLA  : 26
anabaena_p    :  NVYIVGLNQLEDIEIDKVNKPHLPLA  : 26
Synechocys    :  NVYIVGLNQLWDVDIDRINKPNLPLA  : 26
                 N6YIVGLNQL D6 ID46NKP LPLA
```

Motif III

```
                          *
Corn          :  IALFKDIPDIEGDRIF  : 16
Wheat         :  IALFKDIPDIEGDRIF  : 16
Leek          :  IALFKDIPDIDGDKIF  : 16
soy-ppt2      :  IALFKDIPDIEGDKVF  : 16
soy-ppt1      :  IALFKDIPDIEGDKVF  : 16
Cuphea        :  IALFKDIPDIEGDKIF  : 16
Arabidopsi    :  IALFKDIPDIEGDKIF  : 16
Nostoc_ppt    :  IAIFKDIPDIEGDRLY  : 16
anabaena_p    :  IAIFKDIPDMEGDRLY  : 16
Synechocys    :  IAIFKDVPDMEGDRQF  : 16
                 IA6FKD6PD6eGD4 5
```

Fig. 35a

Motif IV

```
                          *
Corn          :  YMLIWKLFYAEYLLIP  : 17
Wheat         :  YMLIWRLFYAEYLLIP  : 17
Leek          :  YMLVWKLFYAEYLLIP  : 17
soy-ppt2      :  YMLIWKLFYAEYLLIP  : 17
soy-ppt1      :  YMLIWKLFYAEYLLIP  : 17
Cuphea        :  YMLIWKLFYAEYLLIP  : 17
Arabidopsi    :  CYMLIWKLFYAEYLLLP  : 17
Nostoc_ppt    :  YQLIWKLFFIEYLIFP  : 17
anabaena_p    :  YQLIWKLFFLEYLMFP  : 17
Synechocys    :  YQLIWKLFFLEYLLYP  : 17
                 fY f6W4LF5 EYL6 P
```

Fig. 35b

Motif A

```
                                *
Nostoc_ppt  :  AFWKFSRPHTIIGT  : 14
anabaena_p  :  AFWKFSRPHTIIGT  : 14
Synechocys  :  AFWRFSRPHTIIGT  : 14
Corn_ppt    :  AFYRFSRPHTVIGT  : 14
soy-ppt2    :  AFYRFSRPHTVIGT  : 14
soy-ppt1    :  AFYRFSRPHTVIGT  : 14
Arabidopsi  :  AFYRFSRPHTVIGT  : 14
Cuphea_ppt  :  AFYRFSRPHTIIGT  : 14
               AF54FSRPHT6IGT
```

Motif B

```
                          *         20
Nostoc_ppt  :  NVYIVGLNQLEDVDIDKINKPHLPLA  : 26
anabaena_p  :  NVYIVGLNQLEDIEIDKVNKPHLPLA  : 26
Synechocys  :  NVYIVGLNQLWDVDIDRINKPNLPLA  : 26
Corn_ppt    :  NIYIVGLNQLFDIEIDKVNKPTLPLA  : 26
soy-ppt2    :  NIYIVGLNQLSDVEIDKINKPYLPLA  : 26
soy-ppt1    :  NIYIVGLNQLSDVEIDKINKPYLPLA  : 26
Arabidopsi  :  NIYIVGLNQLSDVEIDKVNKPYLPLA  : 26
Cuphea_ppt  :  NIYIVGLNQLSDIDIDKVNKPYLPLA  : 26
               N6YIVGLNQL D6 ID46NKP LPLA
```

Motif C

```
Nostoc_ppt  :  IAIFKDIPDIEGDRLY  : 16
anabaena_p  :  IAIFKDIPDMEGDRLY  : 16
Synechocys  :  IAIFKDVPDMEGDRQF  : 16
Corn_ppt    :  IALFKDIPDIEGDRLF  : 16
soy-ppt2    :  IALFKDIPDIEGDKVF  : 16
soy-ppt1    :  IALFKDIPDIEGDKVF  : 16
Arabidopsi  :  IALFKDIPDIEGDKLF  : 16
Cuphea_ppt  :  IALFKDIPDIEGDKLF  : 16
               IA6FKD6PD6EGD4 5
```

Motif D

```
                                *
Nostoc_ppt  :  LQ  W L  I L  F  : 17
anabaena_p  :  LQ  W L  L T  F  : 17
Synechocys  :  LQ  W L  L L  Y  : 17
Corn_ppt    :  MK  W L  A Y L I  : 17
soy-ppt2    :  MK  W L  A Y L I  : 17
soy-ppt1    :  MK  W L  A Y L I  : 17
Arabidopsi  :  CYMK  W L  A Y L L  : 17
```

Fig. 36